5-2018

A Study of Personality Perception Evolution in Mechanical Engineering Design Teams

Hallie Elizabeth Stidham
Clemson University, 93hallie@gmail.com

A STUDY OF PERSONALITY PERCEPTION EVOLUTION IN MECHANICAL ENGINEERING DESIGN TEAMS

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Mechanical Engineering

---

by
Hallie Elizabeth Stidham
May 2018

Accepted by:
Joshua D. Summers, Committee Chair
John R. Wagner
Marisa K. Orr

Distinguished Guest: Marissa Shuffler

ABSTRACT

This research explores the role of personality in engineering design teams in two different populations using the Five Factor Model of Personality. Both the self and peer evaluations of personality were collected in equal increments during one semester. After four iterations, the expectation was that the self-evaluations would be stable over time, peer evaluations would show more agreement with each subsequent iteration and that peers would be better able to identify their teammates personality traits. Results show that the self-evaluations were stable and that the peer evaluations do change over time, with an increase in agreement for the Neuroticism factor in one population. Similarly, peers were shown to be better able to evaluate their teammates personality for the factors of Conscientiousness, Openness, and Extraversion over time. A comparison of two populations working on an engineering design project are presented. Extended results, limitations, and future work are also discussed.

DEDICATION

This thesis is dedicated to my family, for their undying support of my endeavors to pursue higher education, even many miles away. To my Mom for reminding me that no decision is permanent and supporting me no matter what. To my Dad for his constant reminders of my potential. To my sisters for making me laugh during trying times and to Joe for always making sure I have a way to get home. Finally, to my grandparents for their moral and financial support and sparking my interest in pursuing higher education.

Fair Winds and Following Seas…

ACKNOWLEDGEMENT

I would like to thank my advisor, Dr. Joshua D. Summers, and my committee members, Dr. John R. Wagner and Dr. Marisa K. Orr, for their support and guidance during this research. Without their help I would not have improved as a researcher in the ways that I have. I would also like to thank Dr. Marissa Shuffler-Porter and her student Michelle for their guidance on all things related to statistical testing of personality data. Finally, I would like to thank the CEDAR lab members, past and present, for always providing me with feedback and ample encouragement along the way. A special thanks to Doug, AP, and James, for their endless support on every research project, conference paper, and journal article that has been completed along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

MBTI – Myers-Briggs Personality Type Indicator

FFM – Five Factor Model

IPIP – International Personality Item Pool

RQ – Research Question

ME4010- First Semester Senior Design at Clemson University

ME4020- Second Semester Senior Design at Clemson University

CI – Creative Inquiry

IT – Iteration

GLM – General Linear Model

IRR- Inter-Rater Reliability

IRA – Inter-Rater Agreement

R_{wg} (SS) – Inter-Rater Agreement for Slightly Skewed Data

R_{wg} (UN) – Inter-Rater Agreement for Uniform Data

σ^2 – Standard Deviation

O – Openness

C – Conscientiousness

E – Extraversion

A – Agreeableness

N – Neuroticism

Chapter One
WHY STUDY TEAMS?

The use of teams in education, industry, and beyond has long been debated. What are the benefits of using teams? Do they really help? How can teams be optimized to perform at their highest level? Recently, teams have emerged as a focus in industry and thus have become a widely debated topic in academia to try and answer these questions.

In academia we hope to have more control of a team, so that we can expose students to experience all aspects of teamwork to better prepare them for future work on teams. In order to do so we need to first identify what we define as a team. In this work, a "team" is defined as a group of two or more people who are working towards a common goal or predefined task in which they are interdependent, and they disband after project termination (Hughes et al., 2011; Borrego et al., 2013; Chiocchio et al., 2012).

This chapter focuses on research done on teams. It starts with a general discussion of research on teams then moves on to engineering teams in industry and engineering teams in academia. The chapter closes with the motivation to study teams and its relation to engineering.

1.1 Research on Teams

There are many different aspects of teams that can be considered during the team formation process. When studying teams in both academia and industry, it is important to be able to classify the differences between the two. A taxonomy was formed for the classification of collaborative design instances that explore the many traits a design team

can have (Ostergaard and Summers, 2009). By using the taxonomy to look at design instances, the hope is that differences between academic and industry teams can be realized. Traits of a team can range from age, ethnic diversity, and gender of the team members to the team's goal, distribution, and type of communication (Ostergaard and Summers, 2009). In industry, teams are often formed out of necessity by people who are available instead of by people who would work well together. In academia there is more control of the team formation process, and by meaningfully selecting teams in academia it can be shown which teams perform the best in a controlled or semi-controlled environment.

Current research on teams focuses on the team formation process and how it can be improved, and also how teamwork can be assessed for improvement (Hughes and Jones, 2011). This is of interest in many different fields such as engineering, psychology, and computer science. The ability to be a functioning member of a team has been listed as one of the most important traits that employers look for in new graduates (Goldberg, 2010). In engineering, a consistent method for forming teams has yet to emerge in either industry or academia. This goal of this work is to begin to lay a foundation for more systematically forming design teams while considering personality.

1.2 Engineering Teams in Industry

Teams are used in many applications in engineering, including product development teams, design teams, and different types of student teams. In industry, teams can be used throughout the typical product development process from the early stages of problem definition and concept generation to later on in idea evaluation and design reviews (Kratzer

et al., 2010; May and Carter, 2001; Holler et al., 2017; Ostergaard and Summers, 2009; Pahl et al., 1996; Ullman, 2010; Dym and Little, 1999). Industry teams can also be used to create new processes and to make improvements to existing products and infrastructure (Borrego et al., 2013).

There are many instances in which engineering teams in industry have been studied. In one case study, researchers were trying to determine if experienced designers could generate more new concepts by learning and using a new tool (Yilmaz et al., 2013). Another case study looked at team effectiveness based on the organizational context of engineering work teams and found that there were direct relationships between team member satisfaction and the organizational context variables (Doolen et al., 2006). Similarly, a case study was conducted to recommend best practices for teams that are globally distributed (Chen and Messner, 2010). This small subset of case studies on engineering teams in industry shows the need for a better understanding of how teams function, starting during the team formation process.

It has also been shown that senior level student design teams can be used as a model for industry teams. This allows researchers to understand effectiveness of different methods of team formation on teams in academia before implementing them into the industrial environment (Borrego et al., 2013). This is strengthened by instructors use of industry goals and achievements as part of the team projects in academia, because students are then meeting the same requirements that would be set out for an industry team (Borrego et al., 2013). Thus, it is important to understand how team formation effects student teams

in a controlled environment, because the same methods can be applied to industry teams for comparison and validation of results.

1.3 Engineering Design Teams in Academia

Similar to research done on industry teams, many studies have been conducted to identify different characteristics of teams in academia. In academia there are many different types of teams such as student only teams or professor and student teams (Borrego et al., 2013). These teams can vary by size, number of disciplines represented within the team, and background of the team members (Layton et al., 2010). They can also be found in many different settings such as class projects or multidisciplinary research teams (McNair et al., 2011). Diversity amongst team members personality, age, and experiences has become a focus (Kearney et al., 2009). When the team's goal is outcome based, it is important that teammates can cooperate to reach their common goal.

There are many different methods that are used to form student engineering teams. Current methods of team selection include random assignment, self-selection, instructor selection or computer generation and all have identified benefits and drawbacks. Self-selection can be beneficial due to the initial level of team cohesion and thus lead to higher performance however it has a tendency for the team to be overly homogeneous (Bacon et al., 1999). Random assignment has been used because it is seemingly "fair," with all students having an equal chance to work together, yet it often does not result in balanced teams which can lead to bad team experiences (Bacon et al., 1999). Finally, instructor selection has been used but was only found to be implemented 15% of the time. This may be due to the difficulties in implementation and variability used in instructor assignment

methods (Bacon et al., 1999). Also available is selection based on a set of criteria such as gender, schedule, or background (Ohland et al., 2012; Layton et al., 2010). This method considers different characteristics of students and arranges teams based on what instructors deem most important for the project (Layton et al., 2010).

Teams can also be formed based on the learning outcome for that particular team or assignment. Should a team be formed with the goal of successful completion of a project or with the goal of providing students with a learning platform to develop teamwork skills? While performance might be most important in an industry setting an argument is also made that in academia it is important to teach students how to deal with conflict in a team (Lewis et al., 1998). Due to the many applications of teamwork beyond the classroom, it is important that students are given opportunities to work in different types of teams, allowing them to develop valuable experience early on in their careers.

The role of engineering design teams in education has changed since the adaptation of teamwork into the undergraduate curriculum by ABET (Layton et al., 2010). In the discussion on student learning outcomes, it has been shown that being part of a cooperative group or team can lead to high levels of achievement and greater productivity of the members (Smith, 1995). Furthermore, it has been shown that specifically for engineering teams it is important to incorporate the concept of teamwork early during higher education so that the students understand how to function as part of a team before they are asked to also apply the breadth of their engineering knowledge in a teamwork setting (Lewis et al., 1998). Previous experimental work on teams has focused on performance of the team members and the team as a whole, without looking at the underlying attributes that could

affect such performance (Kichuk and Wiesner, 1997; Ogot and Okudan, 2006). With the preponderance of methods to form student teams, it is clear that the engineering design education community has yet to converge on a method, or even an underlying goal for teams in academia.

1.4 Motivation and Relation to Engineering

The motivation for this work is to look at how team members' perceptions change by first looking at some of the underlying attributes of student's personalities. This can be used to determine whether or not understanding one's peers play a role in team operation. If it is shown that understanding of one's peers is important in a team setting, then the findings can be considered during the team formation process. This can also allow for the introduction of new learning objectives about teamwork, based on the goal of the project. Finally, in future work the perception of peers' personalities can be related to team performance which is of interest to industry teams.

When discussing the team formation process, many questions are raised about its relation to the engineering discipline. Questions such as, is this engineering? Why can this not be left to psychologists? Finally, why do we need more research on teams? By studying engineers with the help of psychologists who specialize in teams, better recommendations can be made about team performance. Without the help of engineers studying engineers it becomes difficult to make recommendations for changes in the teaming process. Further, like any design problem that an engineer encounters, there are different requirements that must be met in order for the project to be successful. This is similar to the team formation process, however the variables in question are team members instead of different

mechanisms. By looking at the characteristics of these design variables (team members), a model can be built for a high performing team. Even though work has been done that tries to understand design teams, the community still has not converged on a consistent method that provides the wanted results every time. This work is a first step in identifying the underlying attributes of these design variables, with the hope that by identifying how they work together the next steps in the process can be identified.

The goal of Chapter Two is to motivate the choice to study the personality aspect of the teaming process. It will discuss different teaming methods that are currently used and motivate why the presented method was chosen over previous methods. Chapter Two also identifies the research questions that drive the presented study. Chapter Three presents a pilot study that was completed in the Summer of 2017 and lessons learned from the pilot that were implemented into the full-length study that was completed in Fall 2017. The participants, timeline, and instrument used are presented in Chapter Four. Chapters Five and Six present the results for each population, and Chapter Seven is a comparison of the results for the populations. Finally, Chapter Eight presents conclusions and Chapter Nine is future work.

Chapter Two
WHY STUDY PERSONALITY?

Though there are many different attributes of a team, personality of the team members has stood out as a variable of interest in team formation processes (LePine et al., 2011). In order to assess team members' personalities, a method for evaluating personality is chosen. This chapter will present research on personality including two prominent measures of personality, the Myers-Briggs Personality Type Indicator and the Five Factor Model of personality. A comparison will be made between the two, and evidence for the choice of the Five Factor Model will be presented. Finally, the personality test that evaluates the Five Factor Model is presented and the research questions are introduced.

2.1 Research on Personality

There are many different personality tests that have been used to create a model for team composition (Licorish et al., 2009). Some of the most popular tests are the Myers-Briggs Personality Type Indicator (MBTI) (Duhe, 2009), Minnesota Multiphasic Personality Inventory (MMPI) (Lilkovych and Sokol), Sixteen Personality Factor Questionnaire (16PF) (Andre et al., 2011), and the Five Factor Model (FFM) (Lugo et al., 2017). Of these personality tests, two have become prominent measures used in engineering team formation, namely the MBTI and the FFM. The MMPI and 16PF are rarely used in engineering, due to their absence in studies on personality in either ASME or The Design Society.

2.1.1 Myers Briggs Type Indicator

One of the most commonly used measures of personality in industry and engineering is the Myers-Briggs Personality Type Indicator (Berthelsdorf and Stone, 2017; Clinebell and Stecher, 2003; Duhe, 2009; Felder et al., 2002; Freiheit, 2014; Kanji et al., 2015; Licorish et al., 2009; Magness and Roslewicz, 2009; Karn and Cowling, 2006; Montequín et al., 2013). This model is based on Jung's theory and categorizes people into one of sixteen different personality profiles (Myers et al., 1998). Based on the survey responses of the individual, they are classified into either extraverted or introverted, sensing or intuitive, thinking or feeling, or judging or perceiving (Myers et al., 1998). These categories are then used to assign a four-letter profile to each individual. For example, someone that it "ISTJ" is introverted (I), sensing (S), thinking (T), and judging (J) (Varvel et al., 2004). A description of each of the classifications can be seen in Figure 2.1.

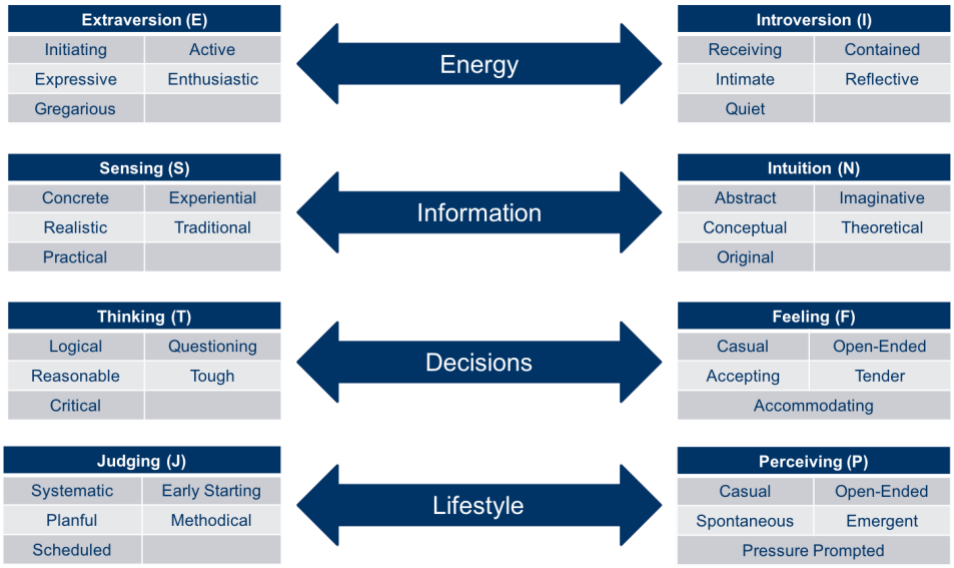


Figure 2.1: MBTI Trait Definitions

In engineering, a prominent method used for engineering design team formation is "Teamology," the process developed by Douglass J. Wilde of Stanford University (Wilde, 2008; Montequín, 2013; Takai and Esterman, 2017; Kress and Schar, 2011). This teaming process uses Jung's theory and the MBTI to assign students to "cognitive groups". Teams are then formed by assigning members with high and low scores in each cognitive group to each team, covering the range of all MBTI scores and forming teams with the greatest cognitive diversity (Wilde, 2008). By using this method, Wilde claims that the teams will all perform at the level that would be awarded a Lincoln Design Award (Wilde, 2008).

The MBTI has been used in many other applications for engineering teams. One case study focused on the validity and predictive ability of the Teamology method proposed by Wilde. The students take the Self-Scorable Form M for the MBTI, which is then fit to the Teamology method to form "cognitive groups". Additionally, the students were able to bid on the projects they found of interest and both factors are combined in the final selection of teams using the Teamology method. This work found that Wilde's factors only accounted for ~10% of variance in design outcomes (Freiheit, 2014).

A similar study proposed a method using the MBTI to help improve team performance by having students take the MBTI and discuss their strengths and weaknesses with their peers (Duhe, 2009). The purposes of the proposed method of teaming are to increase communication, increase the students understanding of others personality traits and develop an appreciation for them, and to try and reduce misunderstandings and frustrations that arise during a team project (Duhe, 2009). There does not appear to be any further work done with this proposed method.

The MBTI has also been used as an avenue to look at student interactions including leadership styles, communication, and organization which can affect the outcome of the project and the team's performance (Magness and Roslewicz, 2009; Montequín et al., 2013; Varvel et al., 2004). Another study used the MBTI as a method for educators to understand their students and thus tailor their teaching styles to the student's personality types (Felder et al., 2002). Similarly, the MBTI has been used to form engineering design teams based on the link between creativity and the different factors of the MBTI such as sensing-intuitive (Shen et al., 2007; Berthelsdorf and Stone, 2017). These are some of the many applications of the MBTI in relation to team formation and studies done on engineering teams.

2.1.2 Five Factor Model

The Five Factor Model of Personality (FFM), also referred to the "Big Five," is another widely used measure of personality. This personality test is preferred by psychologists and has been shown to be able to measure self and peer personality traits (Furnham, 1996). The FFM is composed of five traits, Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Referred to as OCEAN). A high level of openness represents curious, insightful and imaginative people who are original and have wide interests. Conscientious people are organized, reliable, responsible, efficient, and thorough. Having a high level of extraversion leads to assertive, active, outgoing, talkative, energetic, and enthusiastic people. If one is agreeable, they tend to be generous, kind, sympathetic, forgiving, trusting, and appreciative. Finally, if one has a low score on neuroticism they tend to be anxious, tense, touchy, and worrying (John and Srivastava,

1999; McCrae and John, 1992). The meanings of a high and low score for each factor can be seen in Figure 2.2. It is important to note that the FFM does not claim that the five traits that it represents are the only traits that compose one's personality, rather that the five traits represented are the most broad and all-encompassing dimensions of personality (John and Srivastava, 1999; De Raad and Schouwenburg, 1996).

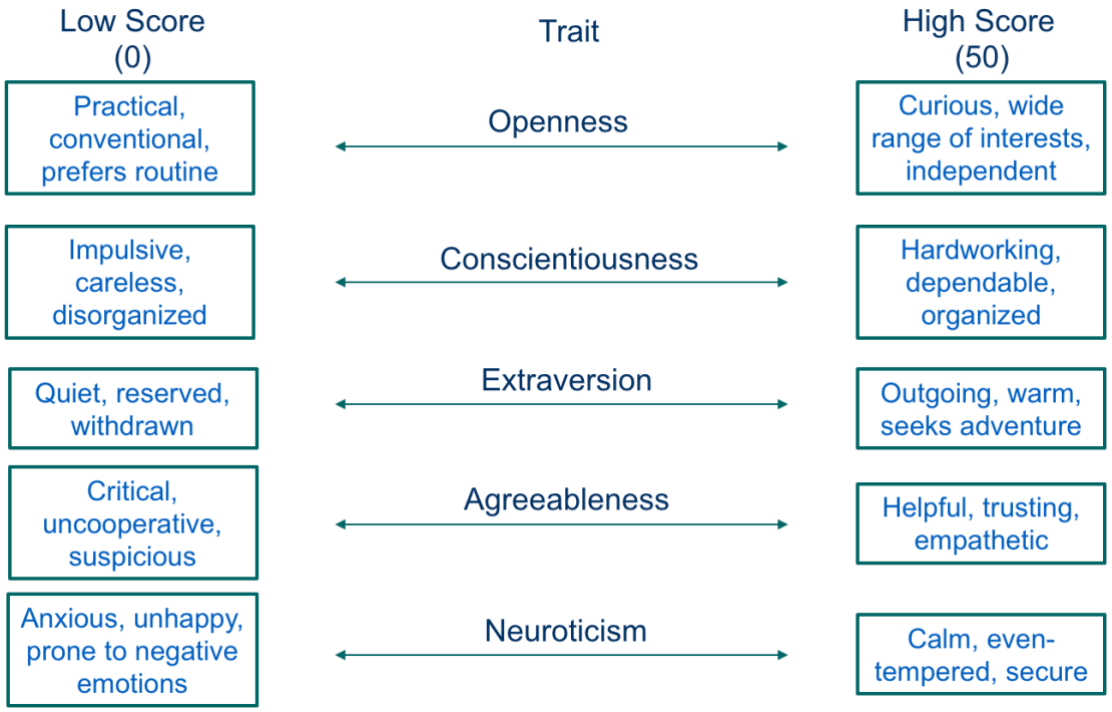


Figure 2.2: Characteristics for each factor in the FFM

Although the Five Factor Model has yet to be widely adopted for use in engineering industry due to the lack of easily distinguishable profiles, it has become more prevalent for use in educational settings investigating engineering design teams. With a need for creating engineering design teams as part of accreditation, instructors have explored many different team formation tools. In one experimental study on engineering team performance related to the FFM, it was found that successful teams showed higher levels of extraversion and

agreeableness with lower levels of neuroticism on the team level (Kichuk and Wiesner, 1997). This experiment was completed in a laboratory setting to control for extraneous factors and thus make the effects of the personality variables of interest more obvious. With 419 subjects, the researchers found that teams that best overcame short term obstacles had lower levels of Neuroticism (Kichuk and Wiesner, 1997). Although these findings are promising, due to the short nature of the project they need to be tested for scalability to a longer-term project.

A similar study on completed research found that for new product development teams, higher levels of agreeableness and conscientiousness on the team level were associated with high performance of the team (Reilly et al., 2002). The researchers define the "team level" as the average value of the factor of interest for the team members. Specifically, they concluded that there is a link between high levels of team openness and project success when creativity is necessary (Reilly et al., 2002).

In interdisciplinary entrepreneurial teams, it was shown that higher performing student teams have high levels of conscientiousness, agreeableness, and openness, as well as a high level of variety in extraversion and neuroticism (Lugo et al., 2017). This exploratory study focused on the team's overall personality as well as the team's personality distribution for each of the five factors. Teams were ranked for performance, and the composition of the top three teams were reported as a suggestion for team composition in future work (Lugo et al., 2017). Though this study was exploratory, the future work on the composition of entrepreneurial teams focused on a design research has implications for other multidisciplinary student teams.

Conscientiousness and Neuroticism were shown to be the best predictors of workplace performance (Hirsh and Peterson, 2008). A Meta-Analysis of job performance in different occupations showed that for all occupational groups studied, conscientiousness was consistently related with high performance (Barrick and Mount, 1991). In fields that involve consistent social interaction, a high score on the extraversion was related to high performance (Barrick and Mount, 1991). These studies show that there is promise for using the Five Factor Model in an industry setting especially when choosing people for positions.

In an academic setting, two studies found that conscientiousness and openness were significantly positively related to course grades (Lounsbury et al., 2003; O'Connor and Paunonen, 2007). In a study on engineering design teams, it was demonstrated that there is potential in using the FFM to help students identify personality strengths and weaknesses as a team and using these traits during an assigned design task (Ogot and Okudan, 2006). Other studies have looked at the creative outcomes from a class assignment, where significant effects of variety of creative outcomes were found for agreeableness and openness and significant effects of novelty were found for conscientiousness, agreeableness, and neuroticism (Okudan et al., 2012; Toh and Miller, 2016).

According to a study on engineering and computer science teams, only 68.1% of teams reported that they achieved communicating clearly with their team members, 66.4% of teams reported they achieved helping one another on the team, and 63.9% of teams reported that they were able to ask for help from other members on their teams (Lingard and Barkataki, 2011). In previous work, it has been shown that team members who understand each other's personalities are better able to accentuate their teammates strengths

and adapt to them, increasing communication between members (Ogot and Okudan, 2006; Clinebell and Stecher, 2003).

2.1.3 Comparing the MBTI and the Five Factor Model

When choosing a personality test for use in this study, two choices were the MBTI and the FFM. Although the MBTI has been widely adopted for use in industry and academia, psychologists warn about flaws in the foundation of this metric (Jackson et al., 1996). It has been shown that this metric is not a reliable or valid measure of personality (Boyle, 1995; Pittenger, 2005; Bjork and Druckman, 1992). This can be attributed in part to the evaluation of personality using a binary scale (Pittenger, 2005). The MBTI categorizes people into one of two categories for each factor and because personality is not binary but continuous, the MBTI often cannot accurately account for ones' personality in its current state. There are also numerous psychometric limitations of this test to consider, including the test re-test reliability. Multiple studies have shown changes in one or more of the attributes for the same subjects with only four to five weeks between administration of the MBTI (McCarley and Carskadon, 1983; Myers et al., 1998).

In contrast to the MBTI, the FFM has been validated in various populations for both self-reports and peer evaluations (McCrae and Costa, 1987; Goldberg, 1992; Digman, 1990). It has also been tested for test re-test reliability and returned favorable results (Goldberg, 1999). Additionally, the FFM has been tested for stability for different age groups and has shown to be stable for the population of interest in this study (Costa and McCrae, 1997). Due to the favorability of the FFM, many attempts have been made to map

the MBTI to the FFM. In two such cases it was found that the types of the MBTI capture multiple factors or do not account for all factors of the FFM (Furnham, 1996; McCrae and Costa, 1989). Specifically, the Neuroticism factor of the FFM influences multiple types of the MBTI and thus prohibits the direct mapping of the MBTI onto the FFM as seen in Figure 2.3.

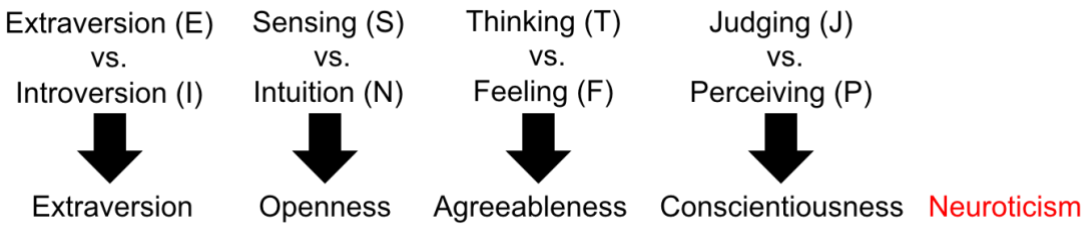


Figure 2.3: Mapping of MBTI onto FFM (Furnham, 1996)

For the purpose of this study, the Five Factor Model of Personality has been chosen as the tool for measuring personality. Due to the nature of the study, the test/retest reliability of the personality metric is important, a trait that the MBTI appears to lack. This work also considers the perception of personality based on the evaluation of one's peers, and while the FFM has been tested for peer evaluation, there is no evidence that the MBTI has been tested for this purpose.

2.2 International Personality Item Pool

The International Personality Item Pool (IPIP) is an open source repository of over 3,000 items and 250 scales that have been developed to look at different aspects of personality. This is an open source resource and as such makes it accessible to researchers for many different applications (Goldberg et al., 2006). This resource was chosen over the

NEO-PI-R, a 240-item test, or the NEO-FFI, a 60-item test, both of which must be purchased for administration of the FFM (Costa and McCrae, 1992).

2.2.1 50 Item IPIP Version of Big Five Markers

In the IPIP, there is a *50 Item version of the Big Five Markers*. This test has been independently tested for reliability and validity on diverse populations and is theoretically grounded in Goldberg's markers for the Five Factor Structure (Goldberg, 1992; Goldberg, 1999; Goldberg et al., 2006; Ehrhart et al., 2008; McCrae and Costa, 1997). The IPIP also provides researchers with a comprehensive set of instructions for use, including how to score the different tests (Goldberg, 1999). This specific survey has been used for multiple studies on personality in engineering and beyond, due to its availability and short time for completion by subjects (Feldt et al., 2008; Kanji et al., 2015; Burton et al., 2010; De Vreede et al., 2012). Another advantage to this survey is that it can be administered using an online form. This provides an advantage in data collection and analysis and allows the subjects to complete the survey at a time that is convenient to them within the identified time frame. The entire 50 Item IPIP version of the Big Five markers with their accompanying factor and key can be found in Appendix A . The first five questions can be found in Table 2.1.

Table 2.1: Selection of questions from the 50 Item IPIP version of the Big Five Markers

Question	Factor	Key
Am the life of the party.	Extraversion	+
Feel little concern for others.	Agreeableness	-
Am always prepared.	Conscientiousness	+
Get stressed out easily.	Neuroticism	-
Have a rich vocabulary.	Openness	+

2.2.2 Scoring of the 50 Item IPIP Version of the Big Five Markers

Each of the fifty questions in the *50 Item Version of the Big Five Markers* correlates to one of the five factors: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The questions are answered using a 5-point Likert scale, and are plus keyed or minus keyed depending on the wording of the question. The number of each type of question and associated factor can be seen in Table 2.2 and an example question can be found in Figure 3.1.

The responses from each of the items are translated to a 50-point scale for each of the five factors. For a plus keyed item, a value of 1 is associated with "Very Inaccurate" on the Likert Scale, and a value of 5 is associated with "Very Accurate," for minus keyed items, the associated values are the opposite. Each factor's score is calculated by summing the numerical values of the answers to the questions associated with the factor, after accounting for which questions are plus and minus keyed. For example, if a participant responded to six plus-keyed agreeableness items with "Moderately Inaccurate" (2), and

four minus keyed items with "Very Accurate" (5) then the total for the agreeableness factor would be 16 (6x2+4x(6-5)=16). The scores can range from 0 to 50 for each factor.

Table 2.2: Plus and minus keyed items for factors in FFM

Factor	Plus Keyed	Minus Keyed
Openness	7	3
Conscientiousness	6	4
Extraversion	5	5
Agreeableness	6	4
Neuroticism	2	8

With the possibility that participants would omit answers on the survey, the average of the responses is used to calculate the contribution of the omitted score. This was done to account for the omission of a question and is a common practice in survey data analysis to prevent the entire data set from being omitted from analysis (Verhoeven, 2016). For example, if a participant omits the response to one of ten questions for the Openness factor, then the average for the nine other questions is used as the score on that question and the aggregate 0 to 50 score is calculated.

2.3 Research Questions

The goal of this study is to explore convergence of personality perceptions on teams using the Five Factor Model. This can be done in two ways, first looking at individual evaluations of personality. Second, to see if peer perceptions change, and finally if they agree with the self-evaluations. For the purpose of this study, convergence is defined as there not being a statistically significant difference at a 95% confidence level between self

and peer ratings. Using the FFM, instructors can start to understand on a fundamental level how student personality perception changes during a design project. Eventually, this can be tied to team performance on such a project. The following research questions have been developed to look at student engineering design teams:

RQ1. Over time, will student self-evaluations change?

RQ2. Over time, will student peer evaluations of the same student converge?

RQ3. Will student peer evaluations converge to match student self-evaluations?

For RQ1, it is expected that the student self-evaluations will not change. This is due to the reported stability of the FFM over time. The expectation for RQ2 is that student peer evaluations will converge during the course of a design project, within the peer perceptions of the ratings. For RQ3, the expectation is twofold. First, that the students would not be able to identify their peer's personalities based on an initial meeting, not necessarily having worked together previously. Next, that through continued shared experiences, the students would be better able to assess their peer's personalities at the end of the project, which can also be seen in RQ2. The relationships between the self-evaluations, peer evaluations, and iterations of the FFM survey can be seen in Figure 2.4.

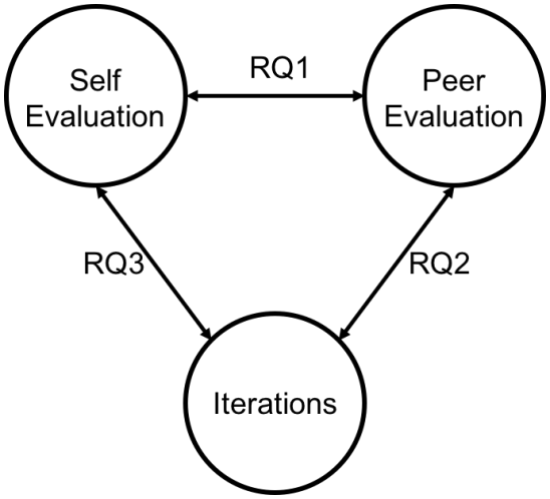


Figure 2.4: Relationship between research questions and survey administration

Chapter Three
PILOT STUDY

In order to identify challenges of studying personality amongst teams in engineering design, a short pilot study was run during the Summer of 2017. The goal of the pilot study was to identify challenges when studying engineering capstone students, and to make changes to address the challenges to ensure a smooth full-length study in the Fall of 2017. The pilot study allowed for preliminary data collections, modification to methods, and calibration of the collection instruments to ensure a more robust design.

This chapter will discuss the pilot study design including a background on the course chosen for the study, the student population used, and the method for data collection. Finally, the results and discussion are presented with an emphasis placed on the changes that were made for the full study design.

3.1 Capstone Design

At Clemson University, the capstone course in Mechanical Engineering is split into two independent courses. For the pilot study, the sample of participants was taken from the second course in the capstone sequence, ME4020, during the Summer of 2017. This course is project based, with teams of students working in parallel on several industry sponsored projects. This course has also been used in several prior research studies on topics such as requirement evolution (Summers et al., 2014; Joshi and Summers, 2015), international differences (Morkos et al., 2014), tool use and information loss (Teegavarapu et al., 2008; Joshi and Summers, 2010; Morkos and Summers, 2010), and requirements use (Morkos and Summers, 2013; Joshi et al., 2011). Team sizes range from four to six students

depending on the project and semester the course is taken. Because the pilot study was completed during the summer, the students are required to complete a project of the same caliber during a five-and-a-half-week timeframe. Design reviews are thus conducted every other day, rather than the weekly 30 min session during a regular semester. The instructor and advisors are not authors on this research.

3.2 Student Population

During the Summer 2017 session of ME4020, there were two industry sponsored projects with a total twenty-three students split into four teams. Project A required students to design a down winding device capable of changing a 25lb spool of monofilament into smaller 3lb spools of monofilament. This project had two teams, both comprised of six students. Project B required students to develop a mobile recharging cart for use in a manufacturing setting. This project also had two teams, one of five students and the second of six students. Both projects required students to meet with their sponsors to conduct preliminary, critical, and final design reviews. This student population at Clemson has been studied in several prior research projects (Thimmaiah et al., 2017; Phelan et al., 2016; Fazelpour et al., 2017; Hannah et al., 2012).

All participants in the pilot study were male, and most had completed either an internship or co-op experience during their undergraduate career at Clemson. The students were assigned to their teams by the instructor using factors such as work experience, technical electives, project interest, teammate requests (positive and negative), scheduling, and personality. This method is similar to how other programs also form teams for capstone projects (Paretti et al., 2011; Bacon et al., 1999).

3.3 Study Method

A series of three iterations of the FFM personality survey for the self and peer were collected during the five-and-a-half-week summer session of ME4020. Iteration 1 was during the first class, Iteration 2 was during week three, and Iteration 3 was at the end of week six. Once the surveys were open to the students, they were given three days to evaluate their own and their peers' personalities.

One survey instrument was used to collect each student's individual personality based on the Five Factor Model. The students were asked to complete the self-evaluation survey at every iteration during the study. An excerpt of the self-survey can be seen in Figure 3.1.

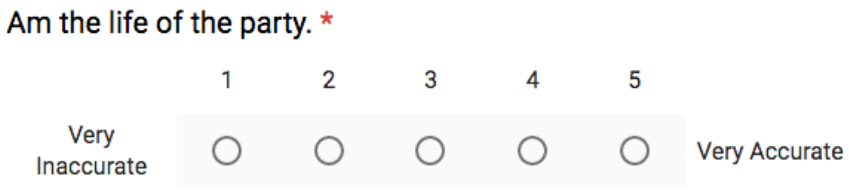


Figure 3.1: Sample Self-Evaluation question and format from pilot study

A second survey instrument was used to collect each student's perceptions of their individual teammates' personalities. The same peer survey was used for each person, thus the students needed to fill out the survey once per person, per iteration. So for example, on a 6 person team one teammate was asked to complete the survey five times for each iteration, for a total of 15 times during the course of the study. A sample questions from the peer evaluation survey used in the pilot study can be seen in Figure 3.2.

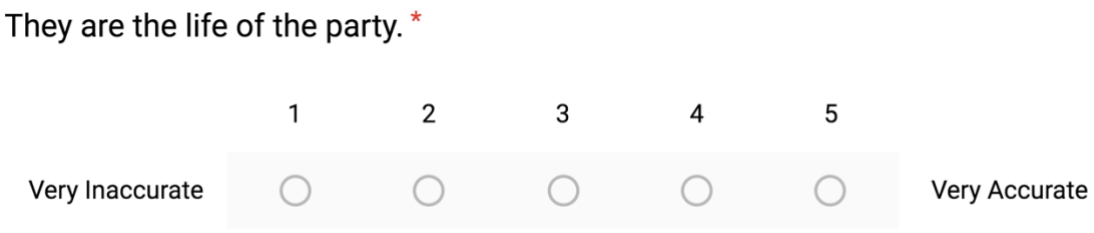


Figure 3.2: Sample Peer Evaluation question and format from pilot study

All surveys were administered using an online form and the students could fill out the surveys at any point during the three-day submission window. The surveys were sent to the students via email and a follow up email was sent on the last day the survey was open for each iteration. Surveys were completed on a voluntary basis, and students agreed to participate in the pilot study at the beginning of the course.

In addition to the FFM questions to evaluate personality, a series of demographic questions accompanied survey. On the self-evaluation survey, students were asked about their co-op or internship experience, and when they had taken the first course in the capstone design sequence. For each peer survey that was completed, the students were asked who they were evaluating, if they had worked with that person previously, and if so, in what capacity they had worked with their peer previously. Because the part of the goal of the study it to look at changes in peers' perception of personality, it is important to capture if teammates knew each other prior to the project.

3.4 Results and Discussion

Results of the pilot study showed that changes to the survey administration, including length and timing, needed to be made to ensure a successful full-length study. Participation in this study was voluntary and 91% of the class agreed to take part in the study over the

course of the semester. During the three iterations of the FFM survey, there was a large range in response rates for both the self and peer evaluations, despite the initial high agreement rate. The self-survey response rates decreased over time and the peer survey response rates stayed flat over the three iterations. The response rates can be seen in Table 3.1. Since the actual participation rates were much lower than anticipated, numerous logistical challenge arose during data analysis. Due to the low response rates, there was not enough statistical power to draw any conclusions from the data set. For this reason, the pilot study data has been omitted from any statistical analysis.

Table 3.1: Pilot study response rates

	Self-Survey Response Rate	**Peer Survey Response Rate**
Iteration 1	47.8%	18.2%
Iteration 2	30.4%	21.8%
Iteration 3	17.4%	18.2%

In addition to the low response rates, the students were unable to successfully follow the instructions for survey completion set out by the researchers. The students would often evaluate their peers' personalities collectively by only completing one peer survey and indicating that the survey was for the entire team. Similarly, students would fill out the survey for some of their teammates and omit others or forget who they had evaluated and evaluate the same person multiple times during one iteration.

The pilot study process allowed the researchers to learn valuable lessons when studying personality amongst student engineering teams. In order to ensure higher response rates in the full-length study, better directions and a separate survey for each team will be

implemented. This should make the survey administration easier on the students, with the hope that it will provide a more complete data set. Additionally, the survey will be administered along with the required course peer evaluations for the full-length study. Although the survey will have no bearing on course grade, the hope is that by administering it concurrently with a course assignment, students will be more motivated to complete the surveys. Students will also be given a longer period of time to complete the survey due to the extended schedule used in a full semester. This allows the students to have more flexibility when completing the surveys. Additionally, in a sixteen-week semester the surveys will be administered only four times, to ensure no survey fatigue. All of these suggestions will be taken into account during the full-length study design, to ensure success in the future.

Chapter Four
STUDY DESIGN: FULL STUDY

After the pilot study was completed, there were many changes that were made to the study protocol. The design of the full-length study is described, including the changes that were made following the pilot study completion. This study is exempt under the IRB2017-117 reviewed at Clemson University. A sample of the agreement statement included on all surveys can be found in Appendix B .

4.1 Study Participants

Two different populations of students were used for the full-length study to be able to compare different sets of students. Both populations were part of engineering design teams working on a semester long project.

4.1.1 ME4010 Participants

The first population of interest was taken from the first course in the sequence of two courses part of senior design at Clemson University, ME4010. During the Fall of 2017, there were 161 students of which 143 were males and 18 were females. Students typically complete this course in the fall of their senior year, and the course is required for graduation. The goal of this course is to teach the students the iterative design approach and forces them to focus on design improvements to existing products. This was done with three distinct phases during the semester. During the first phase, the students had to identify an existing product and define the problem they wanted to address for that product. Next, the students were asked to create multiple concepts that would solve the problem they had identified. Finally, the students chose one of their concepts and build a full prototype which

they had to do a final presentation and report on. These three phases were equally spaced during the semester.

In order for the teams to be determined for this semester long course, the students name two positives and a negative as well as indicate characteristics such as schedule, experience, and qualifications. The options for all positive and negative choices can be teammates with whom they wish or do not wish to work with, or projects that they wish or do not wish to address. After taking into account the student preferences and characteristics, the instructor assigns combinations for all teams trying to make them as balanced as possible. During the Fall 2017 semester there were 26 teams of 6 students and 1 team of 5 students. The instructor of the course is not a collaborator on this research.

4.1.2 Creative Inquiry Participants

The second population of interest was taken from a Creative Inquiry (CI) course within Mechanical Engineering at Clemson University. This CI focuses on advising students as they apply to NASA's Micro-g NExT Program, a design competition put forth by NASA once a year. Since CI's are extracurricular, this course was open to students of different levels and disciplines. The purpose of this course is to allow undergraduate students to focus on a year-long, team based and interdisciplinary research project. Since this course is not required for graduation, the students are self-motivated to be successful.

During Fall 2017, twenty students participated in this CI. The students were split into four teams of five students each and were randomly assigned to have an equal number of students from each academic level represented (sophomores through seniors). The demographics of the CI teams are found in Table 4.1. Some of the different majors

represented included Mechanical Engineering, Materials Science, Computer Engineering, Chemical Engineering, and Bio-Engineering. To apply for the Micro-g NExT Program, the student teams have to design, build, and test a device based on ~15 requirements released by NASA. In 2017, there were four different challenges. Of the four student teams, two chose the "Under Ice Sampling Device" challenge, and the other two chose the "Sharp Edge Detection/Removal Device" challenge. Teams chose their project based on interest, scope (# requirements), and feasibility (# concepts). All four teams wrote a proposal which was submitted to NASA and the two teams on the Sharp Edge Detection/Removal Device Challenge were selected to participate in the Micro-g NExT Program. In May 2018 these teams will travel to Houston, TX to test their devices in the Neutral Buoyancy Laboratory at Johnson Space Center.

Table 4.1: Creative Inquiry Demographics

	Males	**Females**
Sophomores	6	5
Juniors	4	1
Seniors	3	1
Total	13	7

4.2 Study Timeline

A semester at Clemson University is composed of approximately 16 weeks. During this time, the self and peer FFM surveys were administered at four time points, or approximately every four weeks. At each iteration the students were given 1 week to

complete the survey. This was true for both populations. A schedule of the survey administration can be seen in Table 4.2.

Table 4.2: Study Survey Administration Schedule

Week #	**Survey Schedule**
Week 1	Iteration 1 Released
Week 2	Iteration 1 Due
Week 3	
Week 4	
Week 5	
Week 6	Iteration 2 Released
Week 7	Iteration 2 Due
Week 8	
Week 9	
Week 10	
Week 11	Iteration 3 Released
Week 12	Iteration 3 Due
Week 13	
Week 14	
Week 15	
Week 16	Iteration 4 Released & Due

4.3 Survey Instrument

Different survey instruments were created for use in ME4010 and CI. The *50 Item IPIP version of the Big Five Markers* was used for all students and was administered using an online form. For the self-evaluations, students were asked to complete an online form that asked them only about themselves. There was a separate self-survey for ME4010 and

CI. A sample questions from the self-survey can be seen in Figure 4.1. In addition to the FFM questions, a series of demographic questions were asked on the self-survey. These questions included co-op or internship experience and gender.

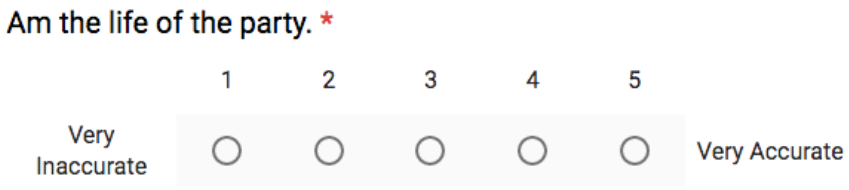


Figure 4.1: Sample self-survey question

For the peer evaluations, a separate online form was created for each team in both ME4010 and CI. This allowed students to evaluate all of their peers' personalities using one survey instrument. This method was chosen to try and eliminate some of the confusion that arose during the pilot study. For each question, a grid was provided to assess the teammates personalities. The students were asked not to complete the evaluation for themselves on the peer evaluation form because the survey instrument has not been externally validated for this application. In the event that students did evaluate their own personality on the peer form, the data was omitted. A sample questions from the peer evaluation form can be seen in Figure 4.2.

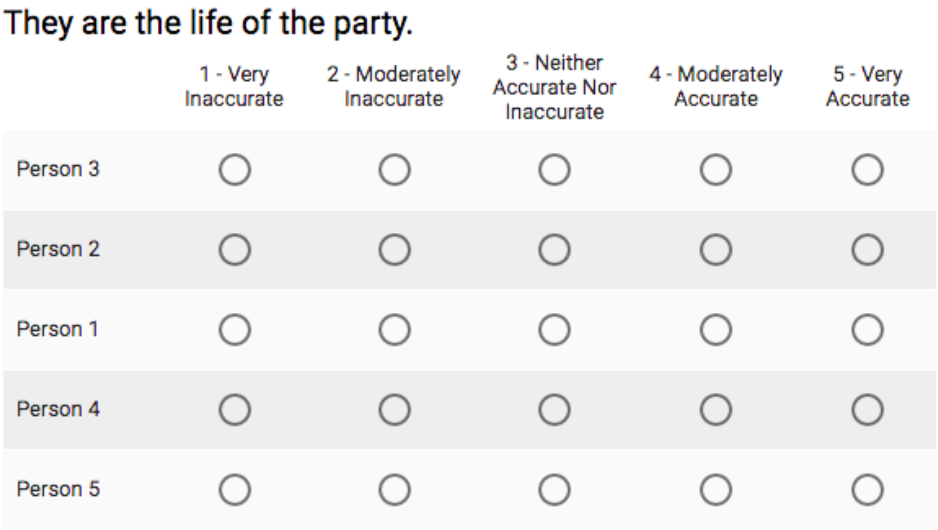


Figure 4.2: Sample peer evaluation question

In addition to the FFM questions asked on the peer evaluation, students were asked if they had worked with one another previously. If the students indicated that they had worked together they were then asked in what setting including social, class, project (curricular), or project (extra-curricular). This allows for previous relationships to be taken into account, since this could have an effect on the convergence of the peer evaluations.

All participation in this study was voluntary. As such, all students agreed to participate (or not) at the beginning of the semester. In ME4010, the FFM surveys were administered concurrently with the mandatory peer evaluations as part of the course. The responses to the FFM survey had no bearing on the course grade for the students that participated. In CI, the students were assigned a complete/incomplete for each survey completed. This was taken into consideration as part of their participation in the course but did not directly affect their course grade.

Chapter Five
RESULTS: CREATIVE INQUIRY

The purpose of this chapter is to present the results of data collection from the Creative Inquiry population taken during the Fall 2017 semester. A comparison of the self-evaluations over time, peer evaluations over time, and self to peer evaluations over time are presented. Over the course of the semester, there was a 98.75% response rate for both the self and peer surveys. Any missing or incomplete surveys were omitted from the data set.

5.1 Self- Evaluations

During all four iterations of the FFM survey, the students were asked to complete both a peer and a self-evaluation of their personalities. The hypothesis for RQ1 is as follows:

Hypothesis 1: There will be no statistically significant differences between self-evaluations over time.

This is due to the reported stability of the Five Factor Model as well as the test/retest reliability (Goldberg, 1999).

The four self-evaluation data points were compared using a General Linear Model (GLM) for Repeated Measures. For all statistical tests the 24th version of IBM's SPSS Statistics was used. An example of the script used to run the tests for each factor can be found in Appendix B . This model compares the iterations and reports if they show statistically significant differences, with a Wilks' Lambda p-value < 0.05 for a 95% confidence interval. This model was chosen because it analyzes the variance of a group of measures when the same measurement is made multiple times for each subject (Hox, 2010).

It also allows for the analysis of the interactions between factors (Hox, 2010). The Wilks' Lambda significance values for each of the five factors can be seen in Table 5.1.

Table 5.1: Wilks' Lambda values for self-evaluations in Iteration 1

Factor	**Wilks' Lambda Significance (p-value)**
O	0.486
C	0.661
E	0.414
A	0.219
N	0.014*

For the factors O, C, E, and A, there were no statistically significant differences between self-evaluations for all iterations. This indicates that these factors were stable for the Creative Inquiry population. For the Neuroticism factor, the reported Wilks' Lambda significance value was less than 0.05, indicating that there was a statistically significant difference in the self-evaluations for this factor (highlighted cell in Table 5.1). This may be attributed to a number of characteristics of this study. First, the sample size for this statistical test is low (N=20), meaning that the differences are more evident when performing statistical tests. Similarly, the Neuroticism factor has been reported to have the most variability of all the factors (McCrae, 1987; McCrae, 1989). In addition, the significant difference could be due to the age of the participants (John and Srivastava, 1999). Since this population included students that were sophomores, they may still be developing more than their senior peers.

When a two-tailed t-test with a 95% confidence interval for an assumed independent two-sample unequal variance is performed on consecutive pairs of iterations, there are no statistically significant differences found. These tests were run because for the purpose of this study, the interest is between consecutive iterations, not necessarily all pairs of iterations. Between Iteration 1 and 2 reports a p-value of 0.614, Iteration 2 and 3 reports a p-value of 0.491, and Iteration 3 and 4 reports a p-value of 0.540. This indicates for consecutive iterations there are no statistically significant differences, but for a comparison of all iterations there are some differences. The descriptive statistics for all iterations of the self-evaluations can be found in Table 5.3.

For all subsequent comparisons, the self-evaluations for each iteration were used to compare to the peer evaluations, ensuring that the statistically significant differences in the Neuroticism factor do not have an effect on the self to peer evaluations. A comparison of Iteration 1 and Iteration 4 for the self-evaluations can be seen in Figure 5.1. Iterations 1 and 4 were chosen for comparison because they had the largest length of time in between evaluations.

Figure 5.1: Comparison of Self-Evaluations for CI in Iterations 1 and 4

For all factors, it is evident from Figure 5.1 that the average score did not change significantly between the two time points. This is further visible in the standard deviation, which does not show much change from Iteration 1 to Iteration 4. Also note that the p-value for a t-test comparing the means of Iteration 1 and Iteration 4 for the Neuroticism factor resulted in 0.595, showing no statistically significant differences between the iterations.

For further analysis, the mean, range, and standard deviation for Iteration 1 and Iteration 4 are shown in Table 5.2. The mean is the average of all self-scores at the identified iteration in the table. The range represents the spread of the scores for that iteration, which shows how much variability there is in the respective iteration. Finally, the standard deviation helps to identify how much deviation there is from the mean at the time point.

Table 5.2: Mean, range, and standard deviation for self-evaluations in Iteration 1 and Iteration 4

Factor	Iteration 1			Iteration 4		
	Mean	Range (0-50)	Standard Deviation	Mean	Range (0-50)	Standard Deviation
O	38.8	16	4.15	39.0	18	4.65
C	41.2	24	5.91	40.3	27	6.11
E	33.2	33	8.97	32.4	28	7.94
A	42.4	13	3.47	41.8	12	3.65
N	40.1	20	5.85	38.8	21	6.26

To identify the distribution of the self-evaluations for the students in this population, a series of distribution vs. frequency plots were developed. One plot for each of the five factors based on the first iteration can be seen in Figure 5.3 to Figure 5.6.

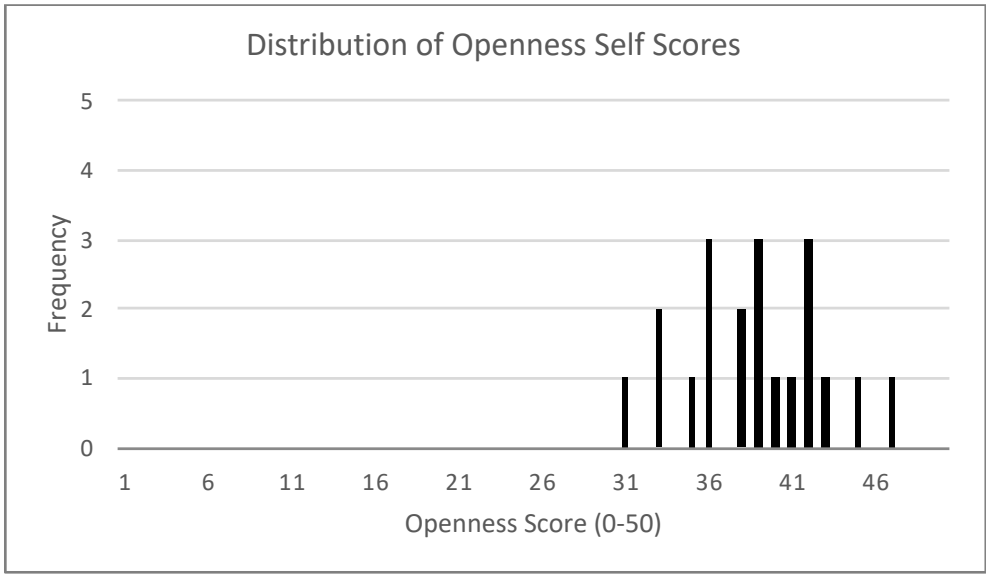


Figure 5.2: CI Distribution for Openness Self-Evaluation Scores

The distribution of the Openness self-evaluations for the CI population shows that as a whole, the subjects are on the high end of the possible score for this factor.

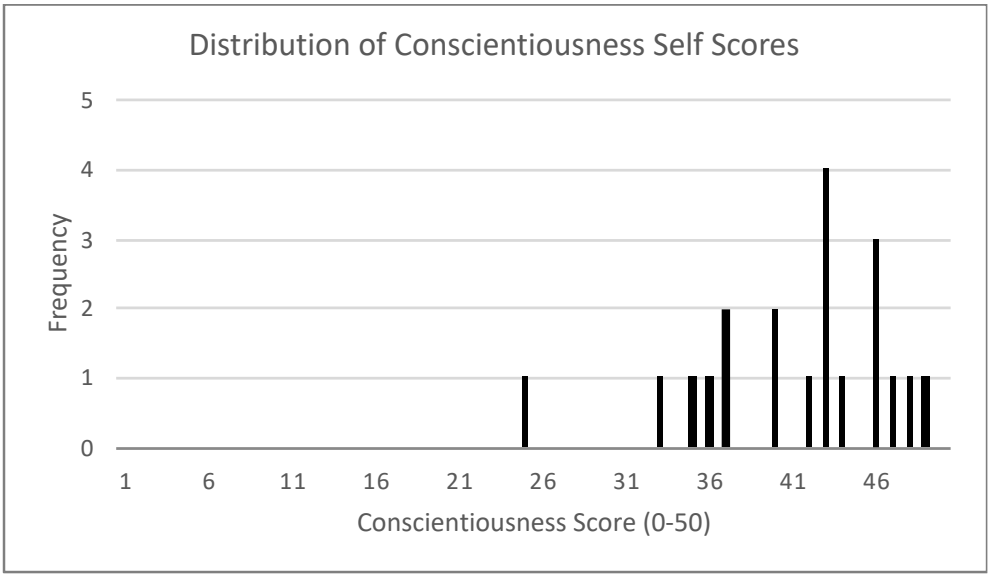


Figure 5.3: CI Distribution of Conscientiousness Self-Evaluation Scores

Distribution for Conscientiousness has a similar pattern to Openness; however, it has a larger range of scores.

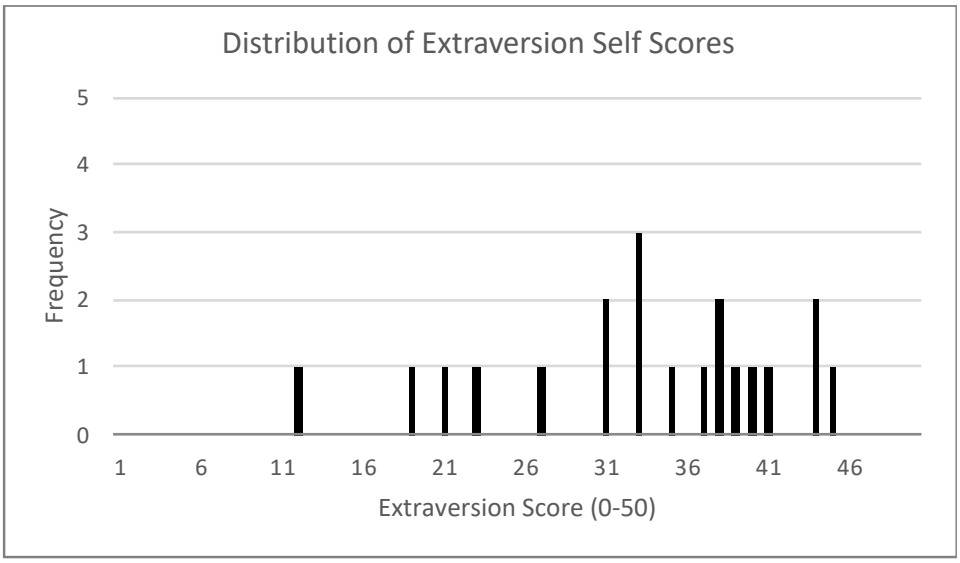


Figure 5.4: CI Distribution of Extraversion Self-Evaluation Scores

Extraversion has the largest range of student self-evaluation scores of any of the five factors. It is the most evenly distributed.

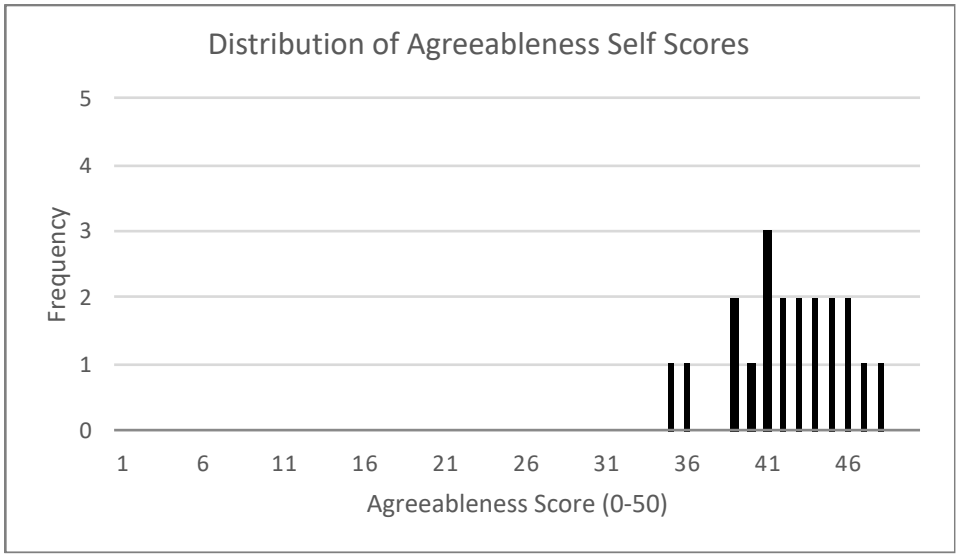


Figure 5.5: CI Distribution of Agreeableness Self-Evaluation Scores

Agreeableness has the smallest range of any of the five factors. Like Openness, the subjects all fall on the above the median of the possible scores.

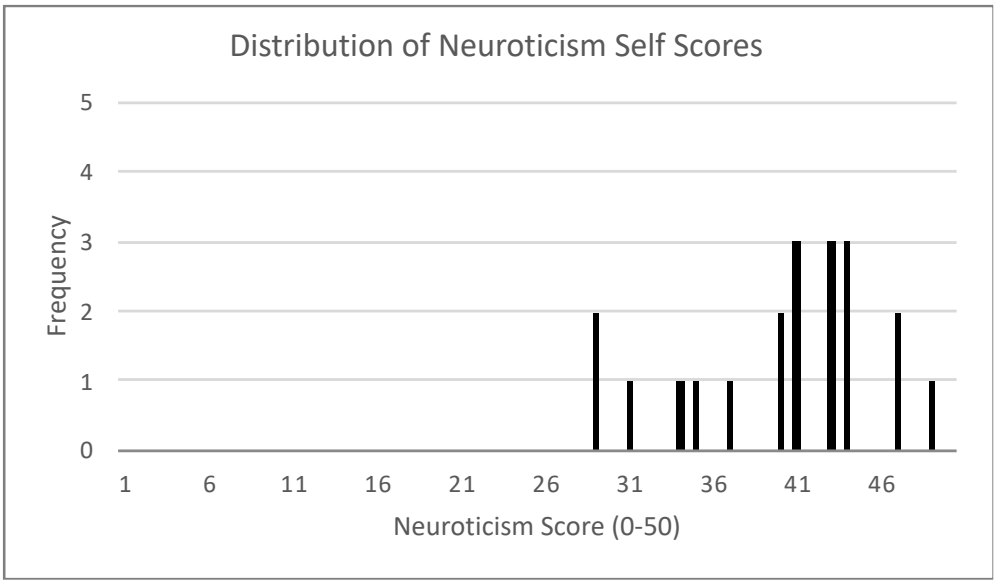


Figure 5.6: CI Distribution of Neuroticism Self-Evaluation Scores

Finally, Neuroticism shows a slightly bi-modal skew of student scores in Iteration 1. For all of the five factors, there is a slight positive skew in the self-evaluations, which can

be seen in Figure 5.2 to Figure 5.6. The factors with the smallest range of self-evaluation scores are Openness (Figure 5.2) and Agreeableness (Figure 5.5). The factor with the largest range in self-evaluation scores was Extraversion (Figure 5.4). The Conscientiousness factor had the highest frequency value for a score of 43, with four students at this score (Figure 5.3). Finally, Neuroticism shows a slightly bi-modal distribution of scores (Figure 5.6). The descriptive statistics for all iterations for each factor have also been listed for consideration in Table 5.3.

Table 5.3: CI Descriptive Statistics for Iterations 1 through 4 of Self Evaluations

Factor		IT1	IT2	IT3	IT4
O	Mean	38.8	38.9	38.3	39.0
	Range	16	16	15	18
	σ^2	4.15	4.67	3.90	4.65
C	Mean	41.2	41.0	40.1	40.3
	Range	24	25	25	27
	σ^2	5.91	5.84	6.32	6.11
E	Mean	33.2	33.7	32.8	32.4
	Range	33	27	26	28
	σ^2	8.97	8.83	7.61	7.94
A	Mean	42.4	41.0	41.1	41.8
	Range	13	19	15	12
	σ^2	3.47	4.04	3.89	3.65
N	Mean	40.1	39.2	37.8	38.8
	Range	20	24	23	21
	σ^2	5.85	5.96	6.31	6.26

The descriptive statistics for each iteration show the variability, or lack there-of, between the self-evaluations. The largest difference between the range of two iterations for all five factors is seven, found between Iteration 1 and Iteration 3 of Extraversion

(highlighted cells). These statistics will also help in the comparison of the self and peer evaluations, found in Section 5.3.

For this analysis, the results of the hypothesis are confirmed. Hypothesis 1 stated that there would be no statistically significant differences between the Iterations over time. Although Neuroticism showed statistically significant differences in the GLM for Repeated Measures, this can be accounted for in the characteristics of the study. Further, the Neuroticism factor failed to reject the null hypothesis for the t-test at a 95% confidence interval for subsequent iterations. For all other factors, O, C, E, and A, the GLM for repeated measures showed that there were statistically significant differences for the Creative Inquiry population.

5.2 Peer to Peer Evaluations

To determine if the peer evaluations converged over time, a series of statistical tests are run to look for statistically significant differences. The hypothesis for RQ2 is as follows:

Hypothesis 2: Through working together on a project the students will understand their peers better over time and thus the peer evaluations will change, converging over time.

Two ways to determine agreement between raters are Inter-Rater Reliability (IRR) and Inter-Rater Agreement (IRA) (LeBreton and Senter, 2008). Inter-Rater Reliability measures if the raters can hit the same spot on a target multiple times for their own ratings. Inter-Rater Agreement measures if the raters can hit the same spot on the target, but not necessarily the middle of the target.

To determine the level of agreement within the iterations, the Inter-Rater Agreement (IRA) was chosen as the method of analysis. This is because for the purpose of the peer evaluations, the hypothesis is that the raters will converge on their agreement over time, without taking into account the true evaluation (self-evaluation). The agreement is calculated using the r_{wg}, which "defines agreement in terms of the proportional reduction in error variance," when multiple raters are assessing a single target using an interval scale of measurement (LeBreton and Senter, 2008). For the raters to agree over time, the r_{wg} value would show an increase with each subsequent iteration with the maximum possible value being 1.

There are two types of r_{wg} that can be calculated, Uniform (UN) and Slightly Skewed (SS) (Meyer et al., 2014). The different r_{wg} values account for either normally distributed or skewed data. Like the self-evaluations, the peer evaluations show a slight positive skew above the norm. For this reason, the slightly skewed r_{wg} values are reported. The uniform r_{wg} values can be found in Appendix B . All r_{wg} values for both samples can be found in Appendix D . The inter-rater agreement for all five factors can be seen in Figure 5.7 to Figure 5.11.

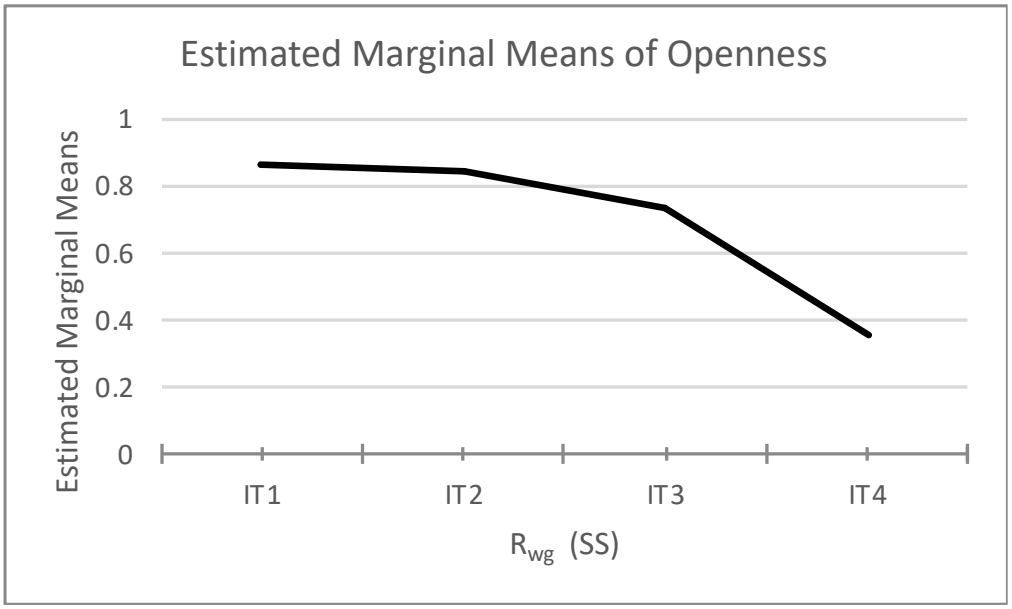


Figure 5.7: CI R_{wg} (SS) Values for Openness Over Time

The r_{wg} for Openness over time decreases at approximately the same rate as Extraversion, seen in Figure 5.9.

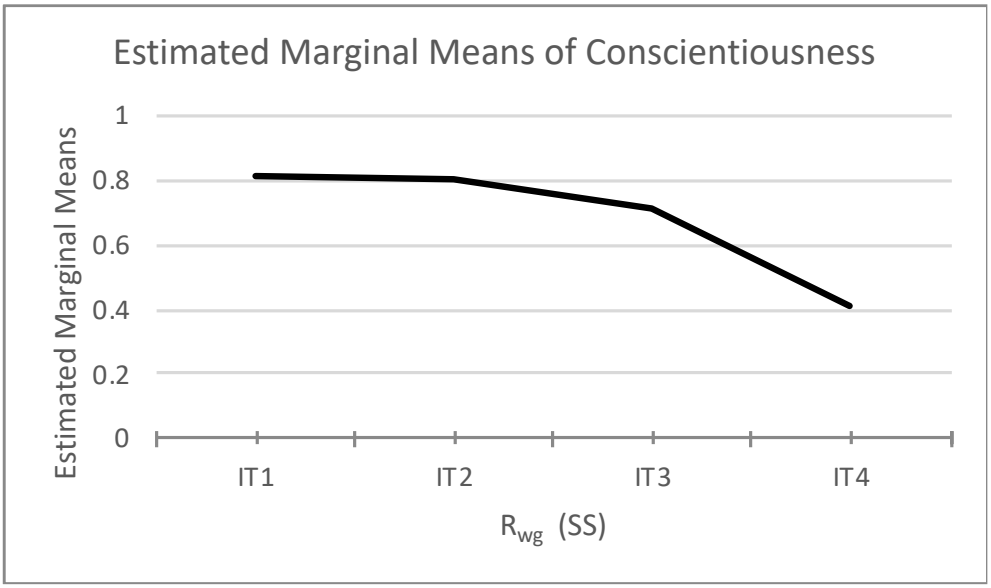


Figure 5.8: CI R_{wg} (SS) Values for Conscientiousness Over Time

Conscientiousness also follows a downward trend, but it does not decrease as much as Openness or Extraversion.

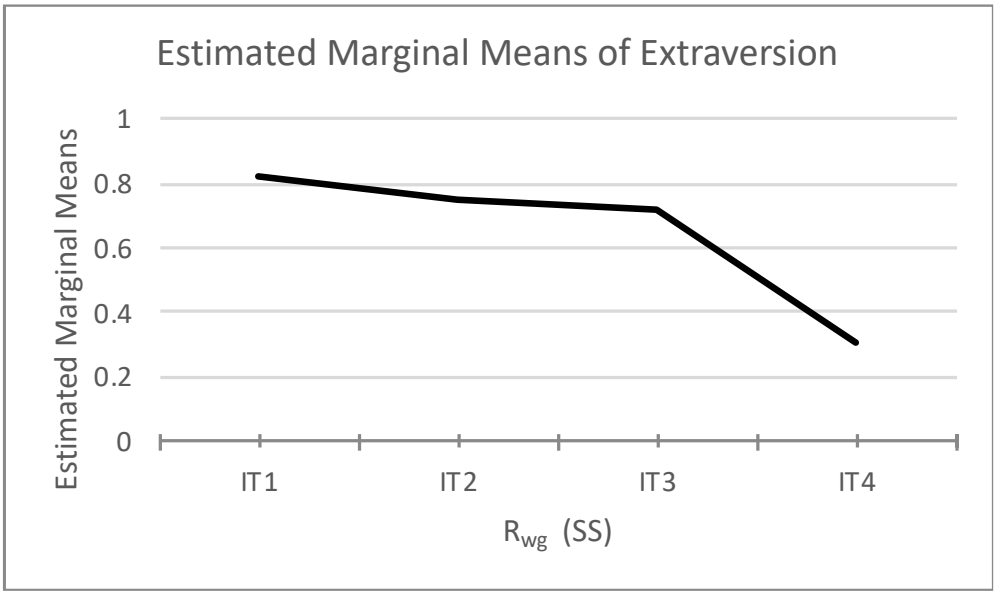


Figure 5.9: CI R_{wg} (SS) Values for Extraversion Over Time

Extraversion has one of the largest drops in IRA from Iteration 3 to Iteration 4.

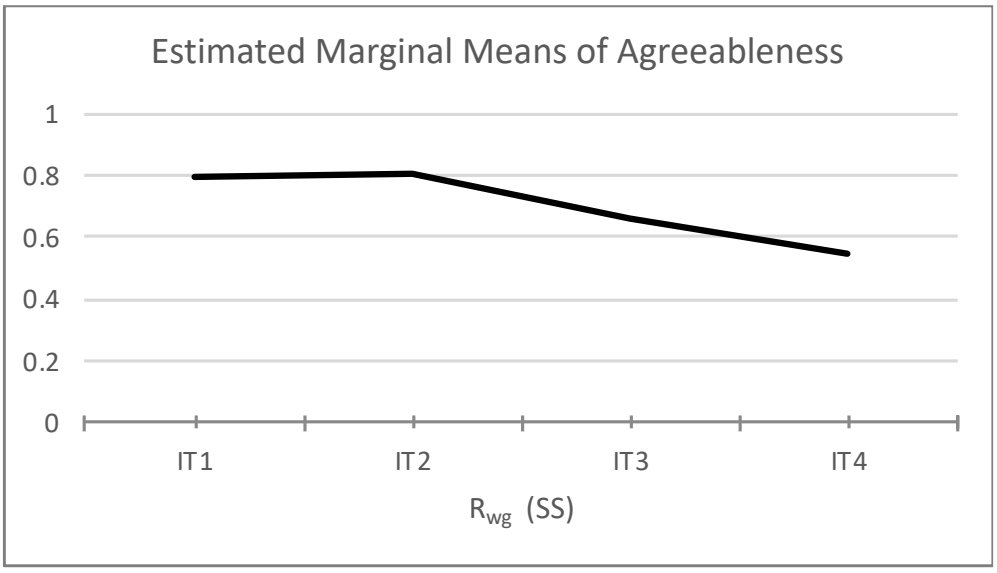


Figure 5.10: CI R_{wg} (SS) Values for Agreeableness Over Time

Agreeableness is the only factor that did not see a decrease in IRA from Iteration 1 to Iteration 2. However, after Iteration 2 the IRA decreases until Iteration 4.

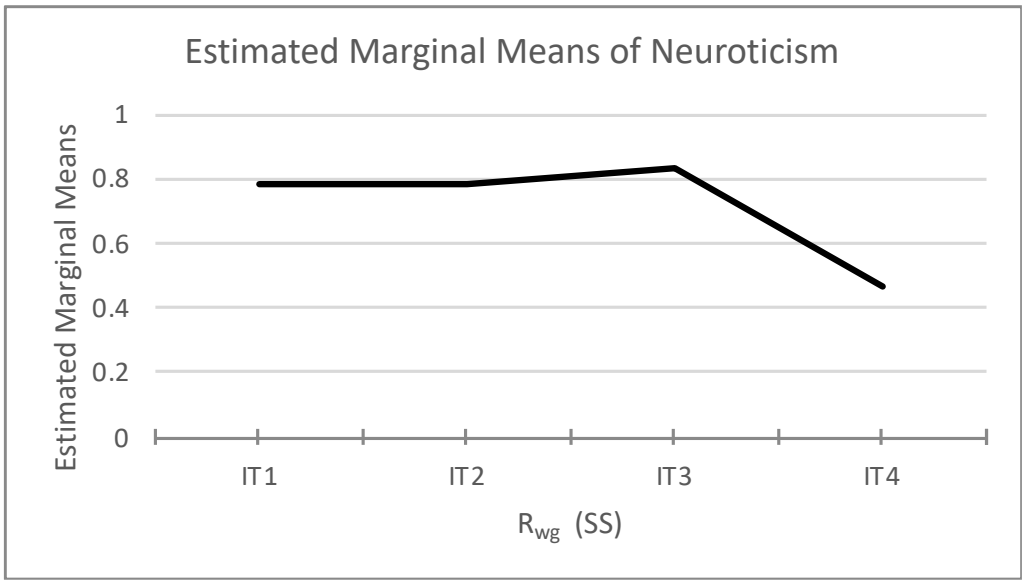


Figure 5.11: CI R_{wg} (SS) Values for Neuroticism Over Time

For Neuroticism, the IRA increases from Iteration 1 to Iteration 3, then decreases from Iteration 3 to Iteration 4. In general, all factors show a downward trend in inter-rater agreement. The factors of Openness (Figure 5.7) and Extraversion (Figure 5.9) saw the greatest decrease between Iteration 1 and Iteration 4, at ~0.5. Extraversion also had the steepest decline between two subsequent iterations, with a drop of 0.4 between Iteration 3 and Iteration 4 (Figure 5.9). Openness and Neuroticism were similar, with a drop of 0.39 and 0.37 both between Iteration 3 and 4, respectively. The Conscientiousness (Figure 5.8) and Agreeableness (Figure 5.10) factors both saw a gradual incline in the IRA over time.

These results show that for the Creative Inquiry population, the peer evaluations were not converging over time (RQ2). This could be due to factors such as small sample size, variety in age of participants, or even differences between different major areas of study.

It could also be due to there being higher distribution of gender than is normally found in engineering, with 65% of the class male and 35% female, compared to the 12% representation of undergraduate females in Mechanical Engineering at Clemson University[1].

To further identify which iterations had statistically significant differences between their respective r_{wg} values, a post-hoc analysis test of within-subject contrasts was run. The results of this test can be seen in Table 5.4.

Table 5.4: CI Peer Evaluation Significance Values for Within-Subject Contrasts

Factor	Iteration Comparison	Significance (p-value)
Openness	IT1 vs. IT2	0.699
	IT2 vs. IT3	0.510
	IT3 vs. IT4	0.221
Conscientiousness	IT1 vs. IT2	0.800
	IT2 vs. IT3	0.585
	IT3 vs. IT4	0.355
Extraversion	IT1 vs. IT2	0.105
	IT2 vs. IT3	0.487
	IT3 vs. IT4	0.067
Agreeableness	IT1 vs. IT2	0.790
	IT2 vs. IT3	0.359
	IT3 vs. IT4	0.634
Neuroticism	IT1 vs. IT2	0.857
	IT2 vs. IT3	0.156
	IT3 vs. IT4	0.153

[1] Office of Institutional Research at Clemson University (www.clemson.edu/ori)

Although there is a steady decline in r_{wg} values for all factors, there were no instances of statistically significant differences between the iterations reported in the post-hoc analysis. In order for the differences to be significant the significance value must be less than 0.05. Extraversion has the lowest significance score out of all factors between Iteration 3 and Iteration 4 with a value of 0.067 (shaded in Table 5.4). However, this score is not low enough to indicate that for Extraversion, the r_{wg} value in Iteration 3 is significantly different from that in Iteration 4.

Finally, the mean, range, and standard deviation for each iteration of all five factors are listed in Table 5.5. For the Openness, Agreeableness, and Neuroticism factors, the range increased between each pair of subsequent iterations. This further suggests that the student peer evaluations for these factors were not converging. For the Conscientiousness factor, the smallest range was in Iteration 4, however the range increased between Iteration 1 and 2 and Iteration 2 and 3 before dropping to its smallest value in Iteration 4. This is also the case for the range of the Extraversion factor. The trends for the range of each of the five factors over all four iterations can be seen in Figure 5.12.

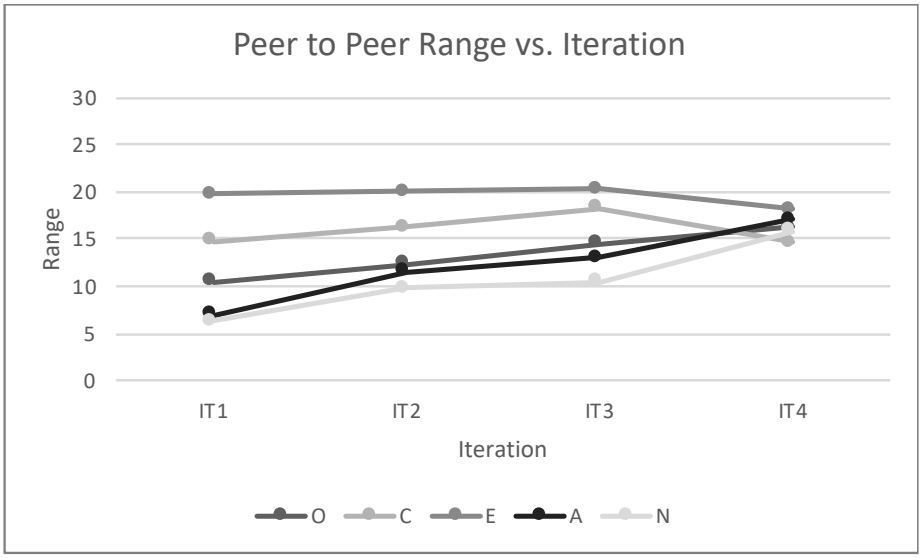


Figure 5.12: CI Trend for Range of Each Factor Over all Iterations

The expectation for the range of the factors is that they will decrease over time. As is evident from Figure 5.12, only the factors Openness and Conscientiousness saw a decrease in the range.

Table 5.5: CI Descriptive Statistics for Peer Evaluation Data

Factor		IT1	IT2	IT3	IT4
O	Mean	39.0	40.7	39.2	38.6
	Range	10.5	12.3	14.5	16.3
	σ^2	2.44	3.20	3.61	3.99
C	Mean	39.0	41.4	39.7	40.3
	Range	14.8	16.3	18.3	14.6
	σ^2	3.60	4.39	4.89	3.72
E	Mean	35.7	38.1	36.6	36.0
	Range	19.8	20.0	20.3	18.2
	σ^2	4.86	4.21	4.74	5.27
A	Mean	38.5	40.6	39.0	39.8
	Range	7.0	11.5	13.0	17.0
	σ^2	1.76	3.18	3.03	4.01
N	Mean	38.6	39.8	39.2	39.3
	Range	6.3	9.8	10.5	15.8
	σ^2	2.01	2.67	2.42	3.52

Unlike the range, there is no general pattern for the standard deviation of the peer evaluations between factors. When looking for peer convergence, a lower standard deviation is preferred because it indicates that most measurements are close to the mean. This helps with the identification of outliers in the data set. An example of this can be seen in Iteration 4 of the Extraversion factor. In this instance the range is 18.2, the lowest for any iteration of this factor, however the standard deviation is 5.27, the highest value for any iteration.

Hypothesis 2 stated that the expectation for RQ2 was that over time, the peer evaluations would converge. The results for this question are inconclusive. Although the r_{wg} values decreased, the range for the factors of Openness and Conscientiousness did decline between some pairs of subsequent iterations. This suggests that it is possible for students to agree on the evaluation of their peer's personality over time.

5.3 Self vs. Peer Evaluations

The expectation for RQ3 is that over time, the students would be better able to assess their peers' personalities having spent more time working together since the initial evaluation. Thus, the hypothesis for RQ3 is as follows:

Hypothesis 3: During the duration of a project, student peer evaluations will converge to student self-evaluations.

To determine whether or not the evaluations were in agreement, the self-evaluation at each respective iteration is compared to the mean peer rating at the corresponding iteration (Zohar, 2010). The mean of the peer ratings was used due to its use in the validation of the instrument across observers (McCrae and Costa, 1987). This is further justified by previous findings that using the mean level for each personality characteristic of the FFM is as consistent in predicting outcomes as the minimum, maximum, or variance in peer evaluations (LePine et al., 2010).

A paired samples test is run between the self-evaluations and the mean peer evaluation with the significance of a two-tailed test with a 95% confidence interval being reported. This test was chosen because it tests whether the means of two variables are equal in the same population. In this case, population one is the self-evaluation and population

two is the mean peer evaluation. The results of the paired samples test can be seen in Table 5.6.

For the factors of Openness, Conscientiousness, and Neuroticism, there were no statistically significant differences reported between the self and mean peer evaluations for any iteration. For the Openness factor, the significance was highest for Iteration 1, then decreased to its lowest point at Iteration 2 before then increasing in Iterations 3 and 4. For the Neuroticism factor, the significance was lower in Iterations 1 and 3 than Iterations 2 and 4. The Conscientiousness factor saw a steady increase in the self and peer evaluations over time, which is what was expected for RQ3. In Iteration 4 for the factor of Conscientiousness, the two samples had a significance value of 0.999, which strongly supports their convergence as the highest possible value for significance is 1. A high score on Conscientiousness represents someone who is hardworking, dependable, and organized while a low score is related to impulsivity, carelessness, and disorganization. In the context of a design project, this factor has significant implications for the team environment which may increase the ability of peers to correctly evaluate this factor better than others over time.

There were three instances of statistically significant differences between the self and mean peer evaluations. During Iteration 2 and Iteration 3 for the Extraversion factor, the self and mean peer evaluations show statistically significant differences. This is also the case for Iteration 1 of the Agreeableness factor. The highest significance value for Extraversion was in Iteration 1 and the highest significance value for Agreeableness was in Iteration 2. All paired samples tests can be seen in Table 5.6.

Table 5.6: CI Paired Samples Test for Self vs. Mean Peer Evaluations

Factor	Iteration	Significance (2-tailed) (p-value)
Openness	IT1	0.811
	IT2	0.133
	IT3	0.495
	IT4	0.722
Conscientiousness	IT1	0.138
	IT2	0.723
	IT3	0.824
	IT4	0.999
Extraversion	IT1	0.215
	IT2	0.038
	IT3	0.041
	IT4	0.073
Agreeableness	IT1	0.000
	IT2	0.712
	IT3	0.100
	IT4	0.112
Neuroticism	IT1	0.294
	IT2	0.662
	IT3	0.379
	IT4	0.755

Although not true for all factors, the Conscientiousness factor shows that it is possible for the mean peer evaluations to converge to the self-evaluations over time. This is proven by the steady increase in significance values between the two samples in each

subsequent iteration. For Openness, Conscientiousness, and Extraversion, when excluding Iteration 1 the convergence was as expected, increasing with each subsequent iteration. Between Iteration 2 and Iteration 4 for Agreeableness the agreement decreased, and the agreement for the Neuroticism factor showed no pattern with each subsequent iteration.

The expectation for RQ3 was that over time the peer evaluations would converge to the self-evaluations for all five factors. For the Creative Inquiry population, the results are inconclusive. When ignoring the first iteration, three of the five factors converged as expected. This shows that it is possible for the students to accurately evaluate their peers' personalities during the course of a design project. Because on average the students evaluated their own personality close to the middle of the possible range for the factors, the high level of agreement in Iteration 1 may be attributed to the students not knowing one another and thus choosing a neutral value on the survey. This would lead to high agreement between the self and peer evaluations, but not because the students understand one another's personalities.

Chapter Six
RESULTS: ME4010

The purpose of this chapter is to present the results of data collected from the population of the first semester of senior design, ME4010, taken during the Fall 2017 semester. Like for the CI population, a comparison of the self-evaluations over time, peer evaluations over time, and self to peer evaluations over time are presented. The response rates for both the self and peer evaluations for all four iterations can be seen in Table 6.1.

Table 6.1: Response Rates for Self and Peer Evaluations of ME4010

	Iteration 1	**Iteration 2**	**Iteration 3**	**Iteration 4**
Self-Responses	99%	58%	84%	95%
Peer Responses	98%	87%	89%	94%

Often, the subjects would omit or forget to answer one to two questions for each evaluation (self or peer). Because of this, the average of the questions answered was used to score each of the five factors on a scale of 0 to 50. This allows for missing data to be accounted for without the need for omitting the entire set of survey results.

6.1 Self-Evaluations

All students in ME4010 were asked to complete both the self and peer evaluation of personality using the Five Factor Model at all four iterations during the semester. For the self-evaluations, the expectation is that they will show no statistically significant differences over time (Hypothesis 1 for RQ1). A General Linear Model for Repeated Measures was run to test for statistically significant differences at a 95% confidence level.

The significance levels are shown with Wilks' Lambda values, as seen in Table 6.2. If the Wilks' Lambda value is less than 0.05, it is indicated that there are statistically significant differences between the self-evaluations, meaning self-evaluations are changing between iterations.

Table 6.2: Wilks' Lambda Values for Self-Evaluations in ME4010

Factor	**Wilks' Lambda Significance (p-value)**
O	0.602
C	0.939
E	0.099
A	0.438
N	0.169

The Wilks' Lambda values indicate that there were no statistically significant differences between the self-evaluation scores over time. Of the four factors, the lowest significance value was for Extraversion as 0.099 (highlighted cell in Table 6.2). Although this value is the lowest, it is still greater than 0.05, which indicates that there are no significant differences. Of all five factors Conscientiousness has the highest significance level, at 0.939 between iterations, showing that this factor is the most stable for this population over time.

The next step in analyzing the self-evaluations is to look at the distribution for each of the five factors. To do so, a series of frequency vs. distribution plots were developed. Using Iteration 1, a plot for the distribution of self-scores for each factor can be seen in

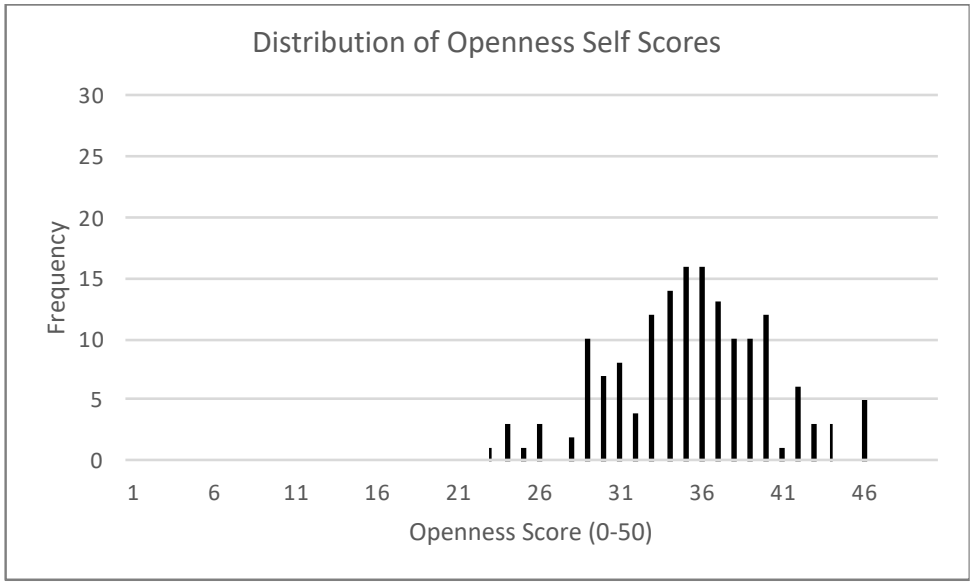


Figure 6.1: ME4010 Distribution for Openness Self-Evaluation Scores

Openness is normally distributed with a slight skew toward the positive side of the median possible points for each factor.

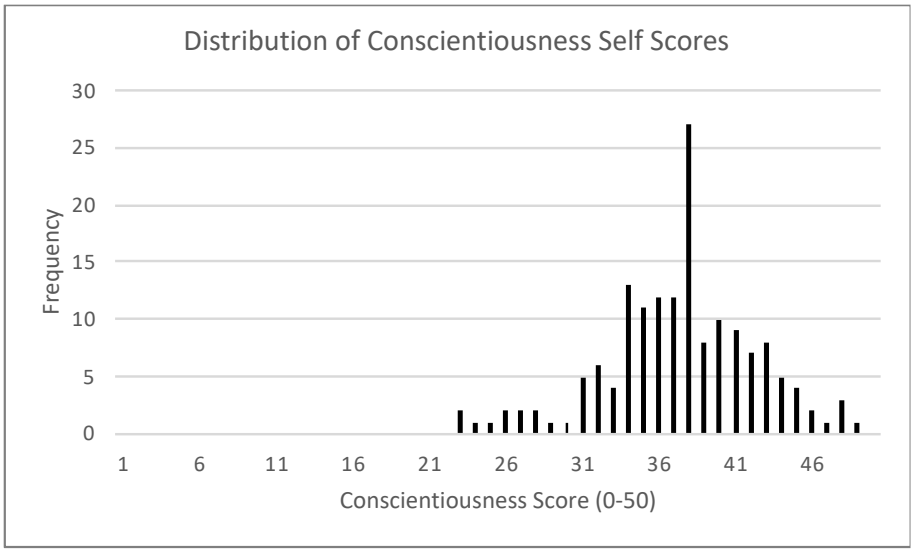


Figure 6.2: ME4010 Distribution for Conscientiousness Self-Evaluation Scores

Conscientiousness has the highest frequency for a single value of any of the five factors with 27 students scoring a 38 out of 50.

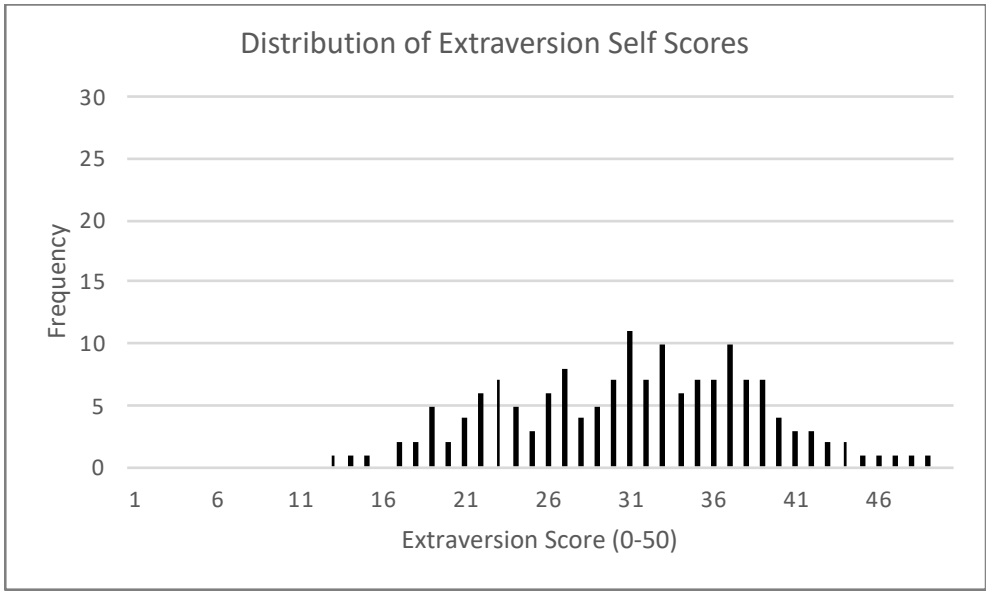


Figure 6.3: ME4010 Distribution for Extraversion Self-Evaluation Scores

Extraversion has the largest range of self-evaluation scores for any of the five factors. It also has the closest mean to the median of the possible scores.

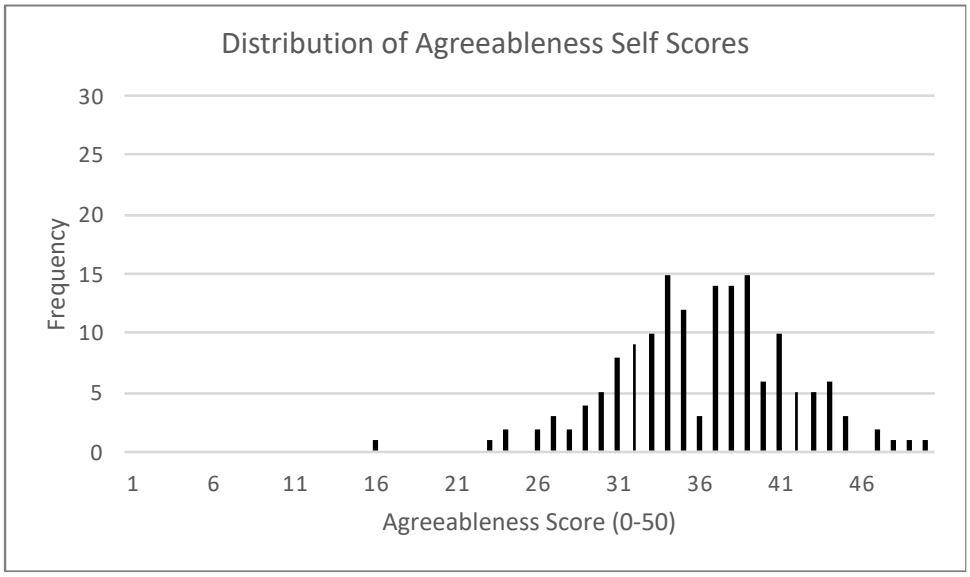


Figure 6.4: ME4010 Distribution for Agreeableness Self-Evaluation Scores

The self-evaluation scores for Agreeableness are normally distributed with a slight positive skew above the median.

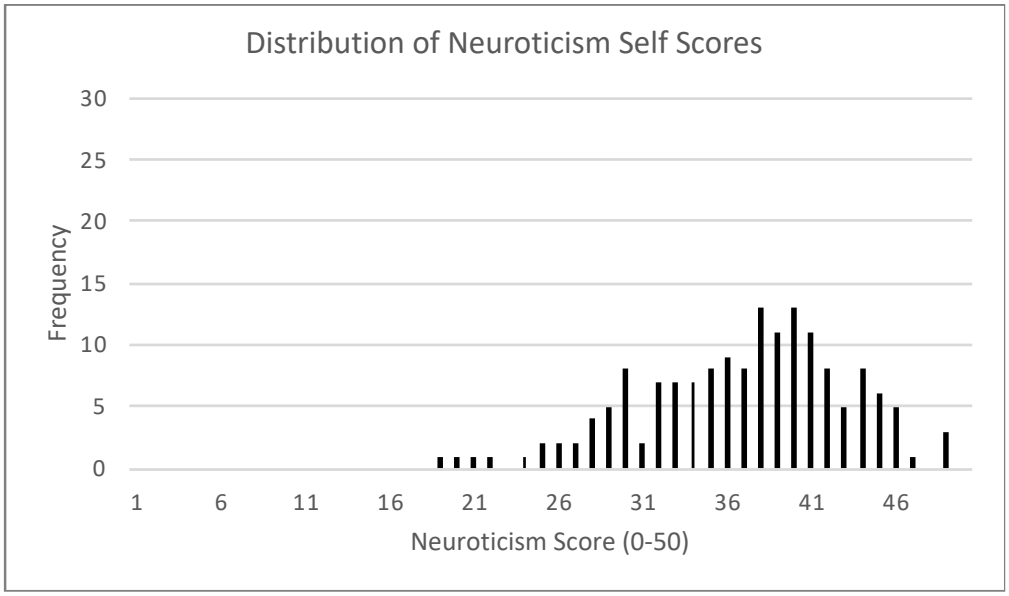


Figure 6.5: Distribution for Neuroticism Self Evaluation Scores

For further analysis, the mean, standard deviation, and range for each factor at each iteration are reported. These values show that in general, the population has a slight positive skew for all factors in regard to the self-evaluations. Extraversion has the largest spread in distribution with Agreeableness and Neuroticism following. The descriptive statistics for all iterations for each factor have also been listed for consideration in Table 6.3.

Table 6.3: Descriptive Statistics for ME4010 Self-Evaluations Over Time

Factor		**IT1 N=160**	**IT2 N=94**	**IT3 N=135**	**IT4 N=153**
O	Mean	35.3	35.7	36.0	36.5
	Range	23	24	25	27
	σ^2	4.81	4.98	4.81	5.33
C	Mean	37.3	38.1	38.2	38.3
	Range	26	31	26	27
	σ^2	5.01	5.98	5.21	5.43
E	Mean	31.1	31.4	32.3	32.7
	Range	36	34	39	39
	σ^2	7.34	7.48	7.60	7.63
A	Mean	36.2	36.6	36.8	37.2
	Range	34	26	24	26
	σ^2	5.41	5.53	5.57	5.41
N	Mean	36.9	37.5	36.8	37.4
	Range	30	28	31	33
	σ^2	6.12	6.14	6.09	6.19

These statistics show the variability in the self-evaluations and emphasize the similarities in ratings between iterations. The largest difference between the range of two iterations is between Iteration 1 and Iteration 3 of the Agreeableness factor, with IT1 having a range of 34 and IT3 having a range of 24. Although this is a relatively large difference, the mean changes only from 36.2 to 36.8 at each time point, respectively.

For RQ1, the Hypothesis is confirmed. The self-evaluations show no statistically-significant differences over time; thus, the Five Factor Model proves to be stable for this population.

6.2 Peer to Peer Evaluations

The next step in the analysis is to determine if the peer evaluations changed over time. For RQ2, the Hypothesis is that the peer evaluations will change and converge over time during a project. In order to test agreement between the peer evaluations, the Inter-Rater Agreement was calculated for all factors over all iterations. The respective r_{wg} value for each set of peer evaluations is reported. The slightly skewed r_{wg} value was chosen due to the slight positive skew on the peer evaluations, similar to the self- evaluations for this population. The uniform r_{wg} values can be found in Appendix F . The Inter-Rater Agreement for all five factors can be found in Figure 6.6 to Figure 6.10.

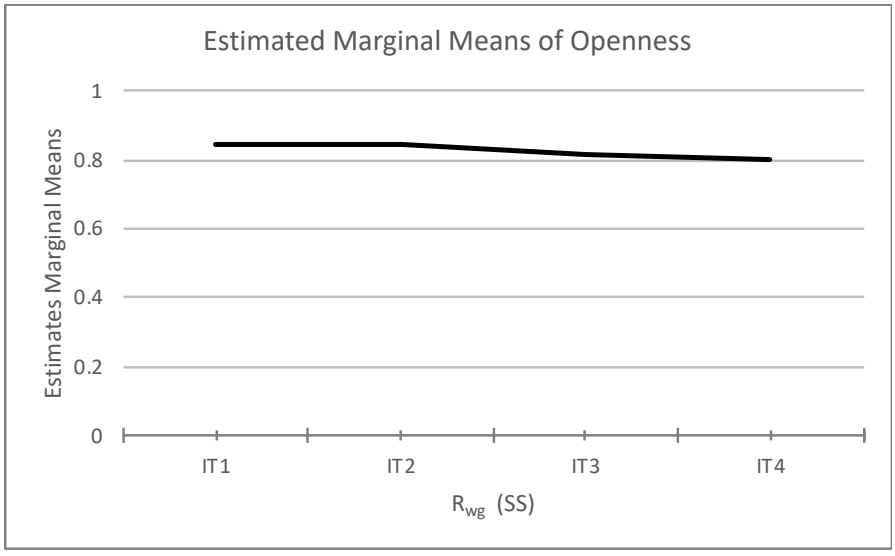


Figure 6.6: ME4010 R_{wg} (SS) Values for Openness Over Time

For Openness, the r_{wg} (ss) values decreased over time, indicating that with each subsequent iteration the students were diverging from one another's ratings of their peers.

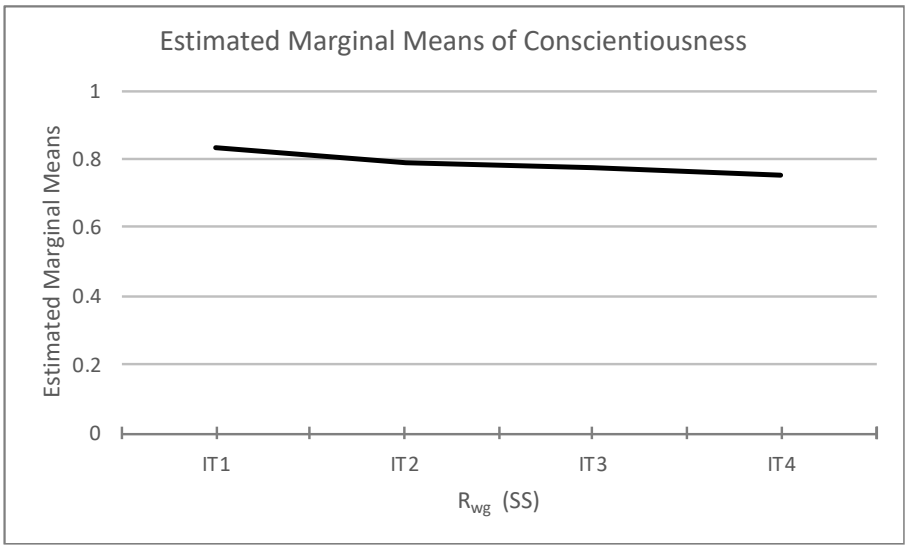


Figure 6.7: ME4010 R_{wg} (SS) Values for Conscientiousness Over Time

The Conscientiousness factor also reflected a decrease in the Inter-Rater Agreement over time.

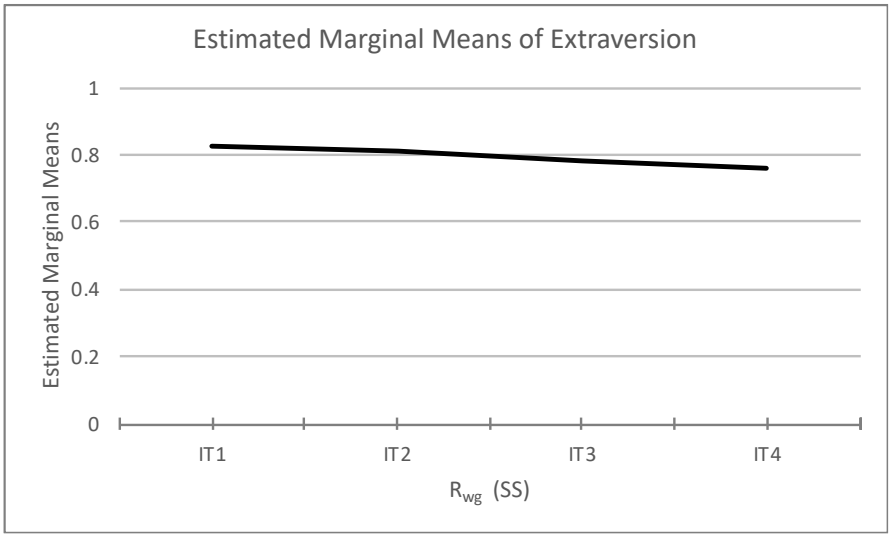


Figure 6.8: ME4010 R_{wg} (SS) Values for Extraversion Over Time

Extraversion saw the most linear decrease in the IRA values over time.

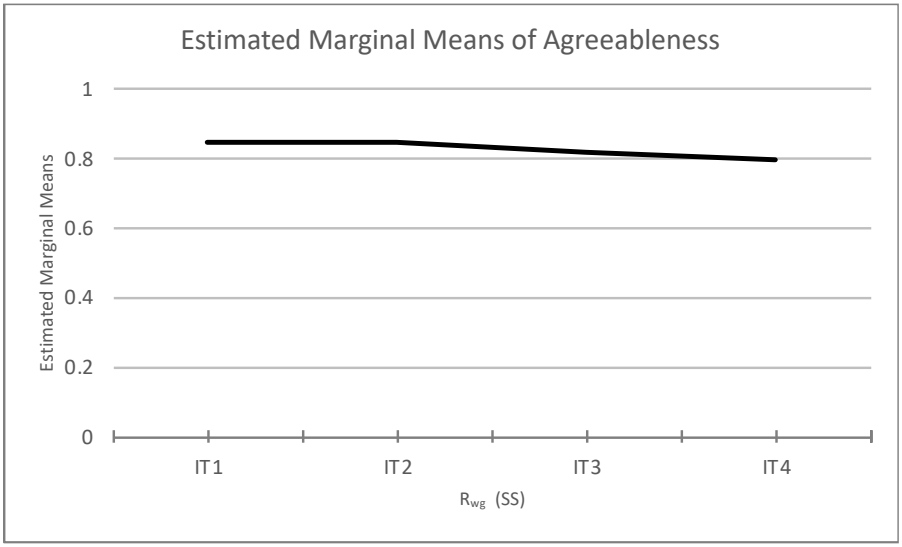


Figure 6.9: ME4010 R_{wg} (SS) Values for Agreeableness Over Time

For the Agreeableness factor, the r_{wg} values stayed constant from Iteration 1 to Iteration 2, then decreased between Iteration 2 and Iteration 4.

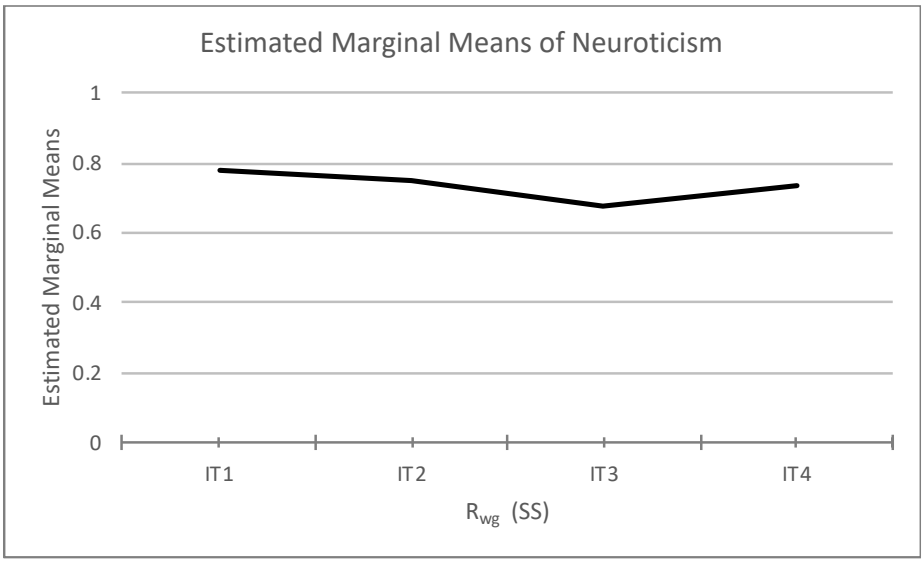


Figure 6.10: ME4010 R_{wg} (SS) Values for Neuroticism Over Time

Finally, the Neuroticism factor showed a decrease in IRA values between Iteration 1 and Iteration 3, then an increase in the inter-rater agreement between Iteration 3 and Iteration 4. In general, there was a decrease in agreement between the raters between Iteration 1 and Iteration 4. Additionally, a test of within-subjects effects is run to identify the significance levels between iterations. For all five factors, the Greenhouse-Geisser significance showed that the r_{wg} values were statistically different from one another. This means that when comparing all possible pairs of r_{wg} values for the peer-evaluations, there are significant differences. This result shows that a post-hoc analysis should be run to look at the subsequent iterations of peer-evaluations to identify if there are statistically significant differences between consecutive iterations. Thus, a post-hoc analysis of within-subject contrasts was run. The results of the post-hoc analysis can be seen in Table 6.4.

Table 6.4: ME4010 Peer Evaluation Significance Values for Within-Subject Contrasts

Factor	Iteration Comparison	Significance (p-value)
Openness	IT1 vs. IT2	0.198
	IT2 vs. IT3	0.003
	IT3 vs. IT4	0.042
Conscientiousness	IT1 vs. IT2	0.005
	IT2 vs. IT3	0.181
	IT3 vs. IT4	0.163
Extraversion	IT1 vs. IT2	0.163
	IT2 vs. IT3	0.101
	IT3 vs. IT4	0.123
Agreeableness	IT1 vs. IT2	0.990
	IT2 vs. IT3	0.043
	IT3 vs. IT4	0.171
Neuroticism	IT1 vs. IT2	0.016
	IT2 vs. IT3	0.043
	IT3 vs. IT4	0.125

From this data, the iterations with significant differences in the r_{wg} values can be identified. All iteration comparisons with statistically significant differences (sig < 0.05) are highlighted in Table 6.4.The Openness factor has statistically significant differences between the r_{wg} values overall, as seen by the Greenhouse-Geisser value of 0.00. However, when taking into account iteration to iteration effects, only between IT2 and IT3, and IT3 and IT4, do the r_{wg} values have statistically significant differences. This is also seen

between IT1 and IT2 for Conscientiousness, IT2 and IT3 for Agreeableness, and IT1 and IT2, IT2 and IT3 for Neuroticism.

In addition to the inter-rater agreement, the range and standard deviation (σ^2) for each iteration is found for the five factors and can be seen in Table 6.5. When analyzing peer evaluations, a lower range and standard deviation are preferred because this indicates an increase in agreement between evaluations. For all factors except Extraversion, the range of peer ratings is larger in Iteration 4 than in Iteration 1. However, for all factors except Neuroticism, the range increased from IT1 in IT2 and IT3, then decreased between IT3 and IT4. The trends in the range for each factor can be seen in Figure 6.11. Similarly, the standard deviation generally increased from IT1 to IT4.

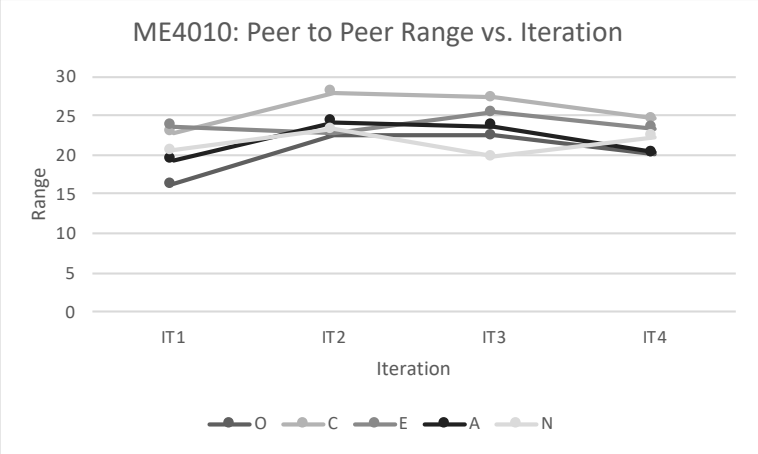


Figure 6.11: ME4010 Trend for Range of Each Factor Over all Iterations

The expectation for the range is that it would decrease over time. When ignoring IT1, this is seen as the trend for the O, C, E, and A factors.

Table 6.5: ME4010 Descriptive Statistics for Peer Evaluation Data

Factor		IT1	IT2	IT3	IT4
O	Mean	35.3	36.4	36.2	36.4
	Range	16.2	22.5	22.4	20.2
	σ^2	2.29	2.71	3.45	3.48
C	Mean	37.3	39.1	38.5	38.6
	Range	22.9	28.0	27.3	24.6
	σ^2	3.20	4.04	4.74	4.19
E	Mean	33.4	34.0	33.9	34.1
	Range	23.6	22.8	25.4	23.4
	σ^2	4.02	4.31	4.28	4.12
A	Mean	36.4	37.6	37.1	37.3
	Range	19.3	24.2	23.6	20.3
	σ^2	2.49	3.16	3.84	3.52
N	Mean	36.1	38.5	37.9	38.9
	Range	20.5	23.3	19.8	22.3
	σ^2	2.93	3.38	4.04	4.02

The hypothesis for RQ2 was such that the peer evaluations would converge over time, indicating agreement within the peers. The results for this question are inconclusive. The peer evaluations did change over time but did not converge in every case. For the factors of O, C, and A, when ignoring IT1, the range decreased between IT2 and IT4, which indicates a greater level of agreement. Thus, while it is possible for peer evaluations to converge, there is not enough information to definitively answer RQ2 at this time.

6.3 Self vs. Peer Evaluations

The hypothesis for RQ3 states that over time, students would be better able to evaluate their peers' personalities having spent more time together, showing a convergence of student peer evaluations to the student self-evaluations. To determine whether the self and peer evaluations were in agreement, the self-evaluation is compared to the mean peer rating at that corresponding iteration. A paired samples test is run between the self and mean peer evaluations with a confidence of 95%. The results of the paired samples test can be seen in Table 6.6.

Table 6.6: ME4010 Results of Paired Samples Test for Self vs. Mean Peer Evaluations

Factor	Iteration	Significance (2-tailed) (p-value)
O	IT1	0.847
	IT2	0.133
	IT3	0.511
	IT4	0.928
C	IT1	0.911
	IT2	0.121
	IT3	0.387
	IT4	0.464
E	IT1	0.002
	IT2	0.019
	IT3	0.021
	IT4	0.034
A	IT1	0.765
	IT2	0.118
	IT3	0.522
	IT4	0.681
N	IT1	0.114
	IT2	0.110
	IT3	0.030
	IT4	0.004

For the factors of Openness, Conscientiousness, and Agreeableness, there are no statistically significant differences found between the self and mean peer ratings in any of the iterations. However, for these factors the significance level starts high at IT1, decreases to its lowest point at IT2, and increases from IT2 to IT4 in all three cases. The only factor

in which the significance level for IT4 is higher than that found at IT1 is the Openness factor.

When looking at the Extraversion factor of self vs. mean peer evaluations, the two data sets are statistically different through all four iterations. Although the significance increases over time, it was never high enough to show that the two data sets were not statistically different. Neuroticism showed a different trend. For IT1 and IT2, the significance level was above 0.05, indicating that the data sets are not significantly different but showing a decrease from IT1 to IT2. At IT3, the significance level dropped below 0.05, indicating that the data sets are statistically different, and it further decreases in IT4.

The results for RQ3 are inconclusive for the ME4010 population. Similar to the peer evaluations, when comparing the self and mean peer evaluations the data show that it is possible for convergence, yet it was not the case in every factor. For the factors of O, C, and A, when IT1 is ignored the data behave as expected. More research is needed in this area for definitive conclusions to be made.

Chapter Seven
RESULTS: COMPARISON OF CI AND ME4010

In this chapter a comparison of the results for each of the research questions will be presented for the two populations. The self-evaluations will be presented first, followed by the peer evaluations then a comparison of the self and mean peer evaluations.

7.1 Self-Evaluations

The expectation for the self-evaluations in both populations was that there would be no statistically significant differences between subsequent iterations due to the reported stability of the Five Factor Model. This was seen in all factors for the ME4010 population, and all factors except for Neuroticism in the Creative Inquiry population. Due to the low sample size in the CI and no statistically significant differences found between subsequent iterations of the self-evaluations when a t-test was performed, the differences in Neuroticism are not of concern. A comparison of the Wilk's Lambda values for the CI and ME4010 populations can be seen in Figure 7.1.

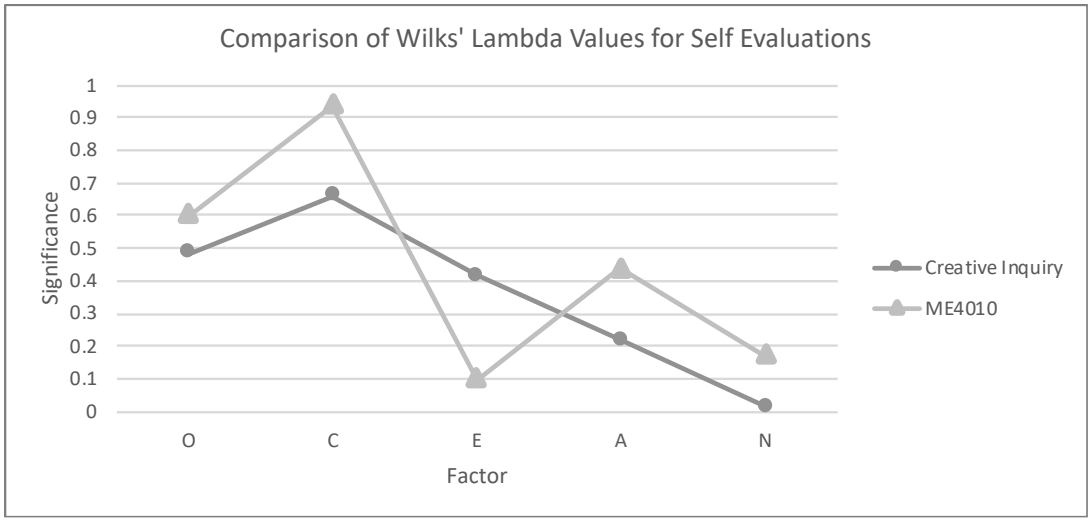


Figure 7.1: Self Evaluation Comparison

For all factors except Extraversion, the ME4010 population showed a higher level of significance when testing for statistically significant differences between the iterations of the self-evaluation data. This could be due to the two different populations and the various stages of development that these students are in. For all other factors, the two populations had the same trend, with Conscientiousness having the largest significance level and Neuroticism the lowest.

7.2 Peer Evaluations

The expectation for the peer evaluations is that the peers would converge over time, showing more agreement on their evaluations of their peers. In both populations the results for RQ2 were inconclusive. The two populations showed that while it was possible for the peers to converge on their evaluations over time, it was not the case for every factor. A comparison of the significance values for within-subject effects can be seen for IT1 vs. IT2, IT2 vs. IT3, and IT3 vs. IT4 in Figure 7.2 through Figure 7.4.

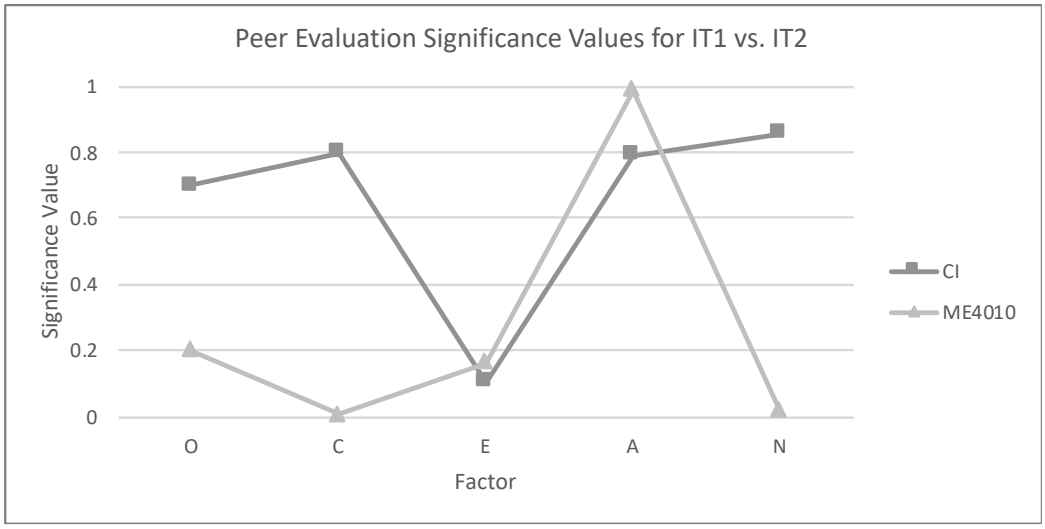


Figure 7.2: Peer Evaluation Comparison for Significance between Iterations 1 and 2

For the comparison of the first iteration to the second iteration, the two populations were in agreement for the Extraversion and Agreeableness factors. For the factors of Openness and Conscientiousness, the two populations showed drastically different significance levels. This was also the case for the Neuroticism factor.

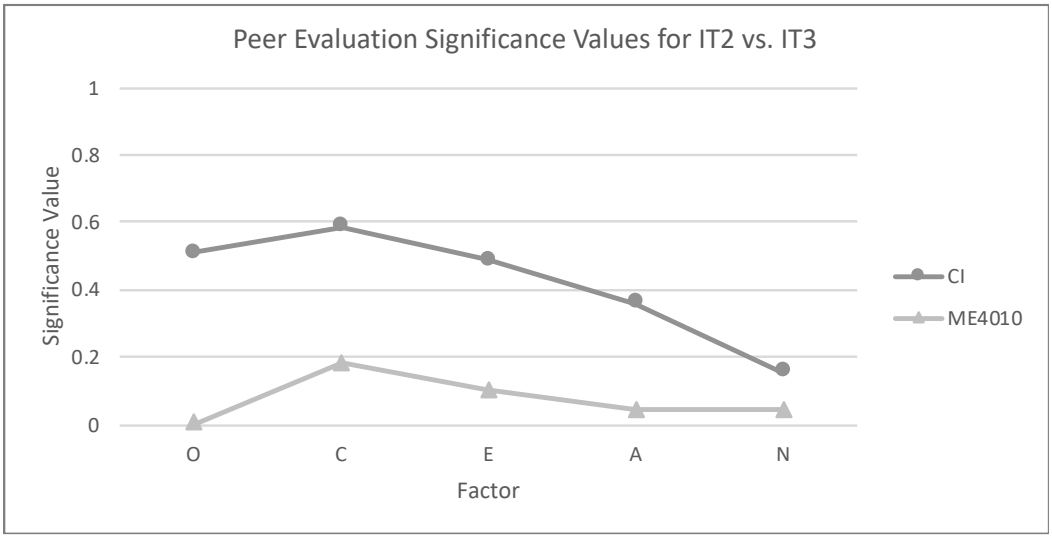


Figure 7.3: Peer Evaluation Comparison for Significance between Iterations 2 and 3

In a comparison of Iteration 2 to Iteration 3 for significance, the two populations showed the same general trend for all five factors.

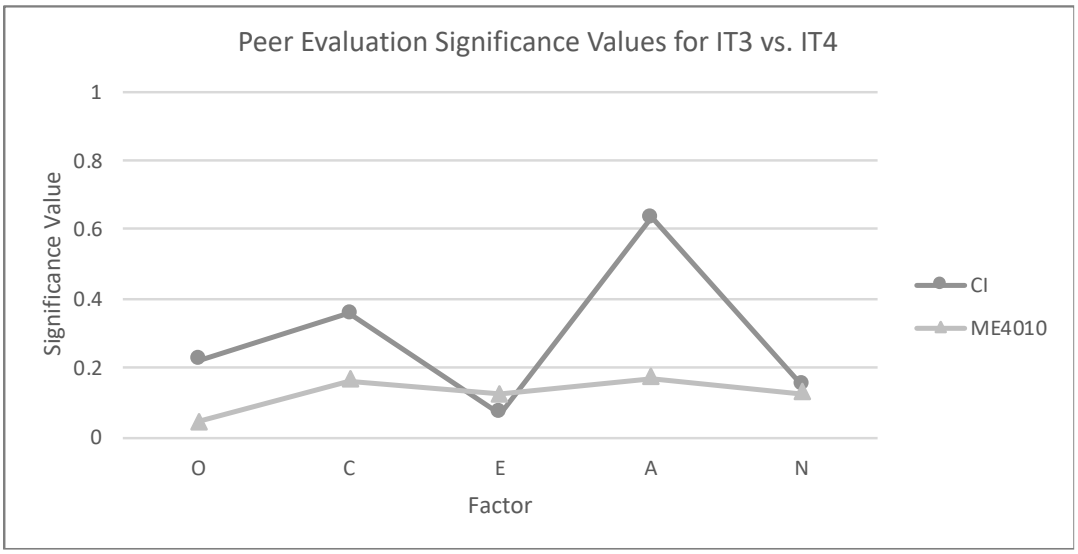


Figure 7.4: Peer Evaluation Comparison for Significance between Iterations 3 and 4

When comparing Iteration 3 to Iteration 4 for significance, the factors of Extraversion and Neuroticism showed the most agreement between populations. In general, the Creative Inquiry subjects showed higher levels of agreement for all iteration comparisons. This may be due to the nature of the project or how much time the students are spending working together. It may also be due to the nature of the project, since the students in CI are on the team as part of an extracurricular project, they may be more self-motivated. Similarly, the CI project is a long-term commitment of ~9 months from project start to finish.

7.3 Self and Peer Evaluation Comparison

The expectation for the comparison of the mean peer and self-evaluations was that over time, the mean peer evaluation would converge to the self-evaluation. In both populations, the results for RQ3 were inconclusive. There were instances in both cases that

the mean peer score did converge to the self-evaluation, however this was not the case for all factors in either population. A comparison of the paired samples test for each iteration between populations can be seen in Figure 7.5 to Figure 7.8.

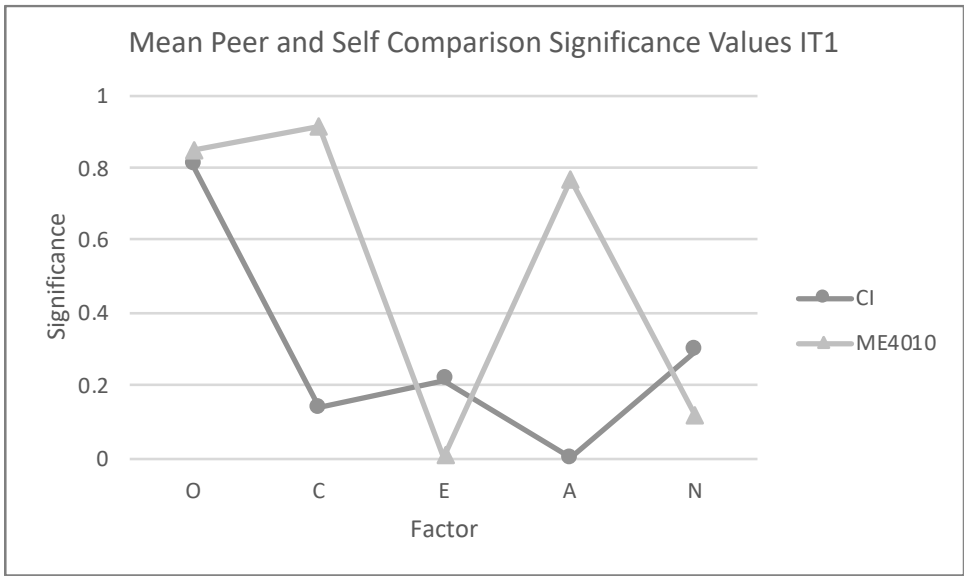


Figure 7.5: Comparison of Mean Peer vs. Self-Evaluation Significance in Iteration 1

In Iteration 1, the Openness, Extraversion, and Neuroticism factors showed the greatest agreement in level of significance between the two populations. For all factors except Extraversion and Neuroticism the ME4010 population showed greater agreement between self and peer evaluations during Iteration 1.

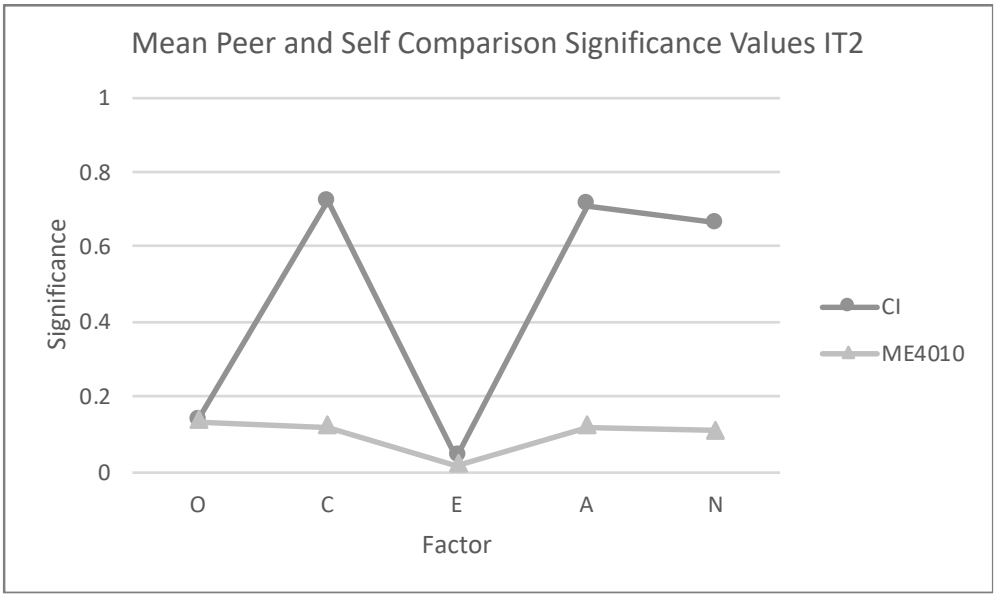


Figure 7.6: Comparison of Mean Peer vs. Self-Evaluation Significance in Iteration 2

During Iteration 2, the CI population showed greater agreement for all factors. Openness had almost exactly the same agreement between the mean peer and self-evaluations of both populations.

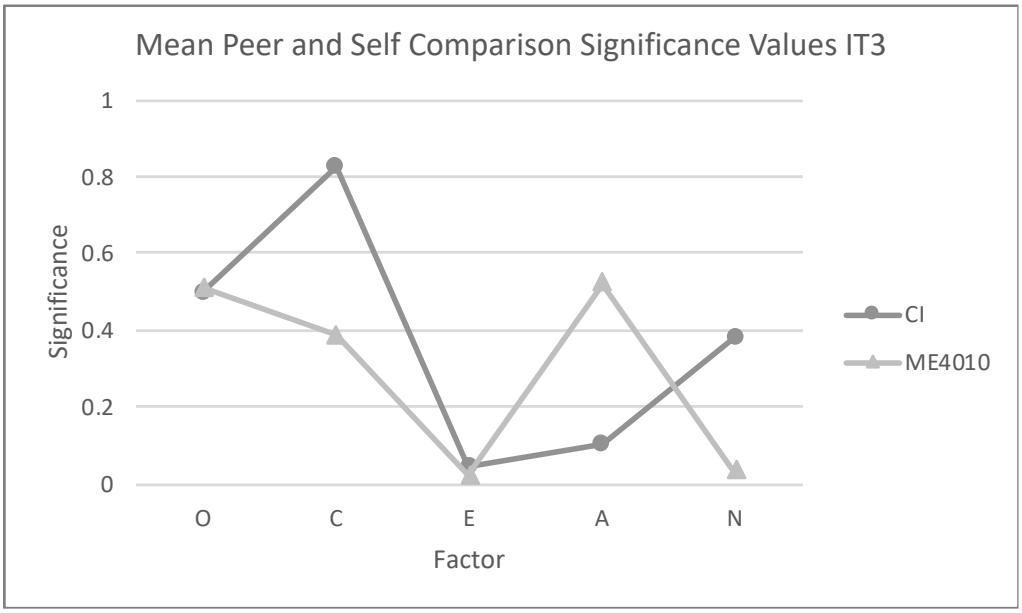


Figure 7.7: Comparison of Mean Peer vs. Self-Evaluation Significance in Iteration 3

Iteration 3 had agreement amongst both populations for the Openness and Extraversion factors. Otherwise, there was greater agreement in the CI population for Conscientiousness and Neuroticism than the ME4010 population.

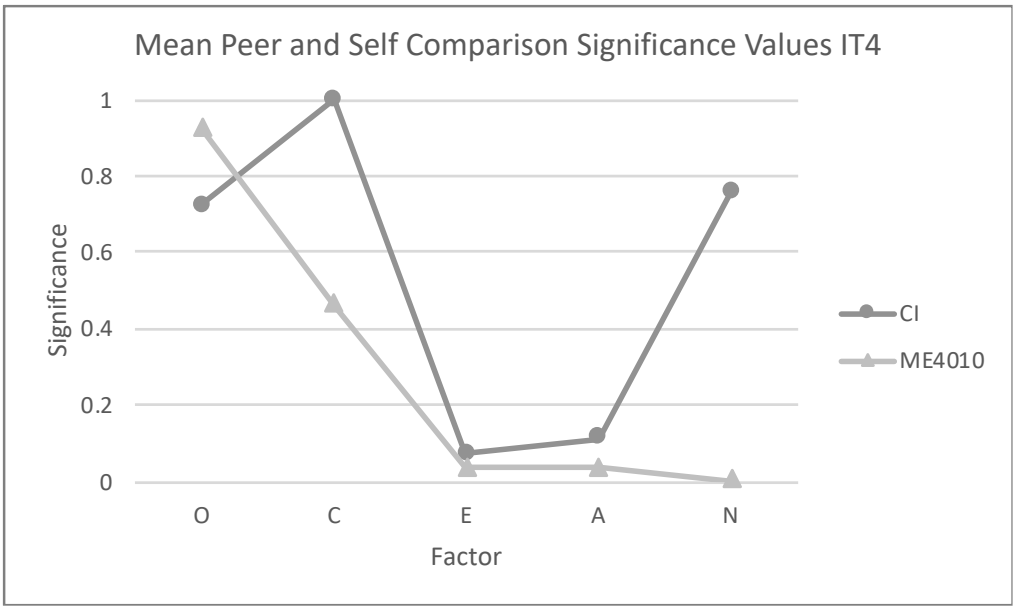


Figure 7.8: Comparison of Mean Peer vs. Self-Evaluations Significance in Iteration 4

In the final iteration the most agreement in significance values between the two populations was found in the Openness, Extraversion, and Agreeableness factors. However, in general there were higher levels of agreement in the CI population in IT4 except for the Openness factor. This may be due in part to the higher agreement amongst the peer evaluations and could also occur due to the nature of the project. Since the design project in CI is more of a commitment and the students are self-motivated to complete the project, it could lead to them trying to understand their teammates better. At the end of the semester in ME4010, it is not likely that the students will work on another project together.

Chapter Eight
CONCLUSIONS

This chapter will present conclusions drawn from this work. It will also list out the limitations of the study.

8.1 Research Question 1

RQ1: Over time, will student self-evaluations change?

Hypothesis 1: Due to reported stability of the Five Factor Model, there will be no statistically significant differences between self-evaluations over time.

Results from both populations show that the self-evaluations will not change over time. This was tested using a General Linear Model for Repeated Measures, which compares iterations and reports if they show statistically significant differences when an evaluation is made multiple times for the same subject. The only instance of statistically significant differences being reported is for the Neuroticism factor in the Creative Inquiry population (N=20). Because there is a low sample size for this population, this may be the source of the significant difference. To account for this the self and peer evaluations for each iteration were compared.

Limitations of the self-evaluations include the time between subsequent iterations and the different personality profiles found in the two populations. Though the survey has been tested for test/retest reliability, it is possible that with more time in between iterations of the survey there would be statistically significant differences. This should be taken into consideration in future work. Similarly, the different self-evaluations found in the two

populations could have an effect on the comparison of self to mean peer evaluations. This limitation will be further explored in subsequent sections.

8.2 Research Question 2

RQ2: Over time, will student peer evaluations of the same student converge?

Hypothesis 2: Through working together on a project the students will understand their peers better over time and thus the peer evaluations will change, converging over time.

Results for this research question are inconclusive. Both populations showed that it is possible for peer evaluations to converge over time, but it was not seen in every case. This was tested by calculating the Inter-Rater Agreement between raters at each iteration. In general, there was a decrease in the range of peer evaluations between Iteration 3 and Iteration 4. Similarly, if the first iteration is not taken into account then the range of peer evaluations tend to decrease over time.

The limitations of the peer evaluations are twofold. First, if students did not know one another and chose the most neutral choice on the Likert scale for the first iteration there was the potential for a high agreement to be reported due to the distribution of the self-evaluations. This can be seen when comparing the average self-evaluation to the mean peer evaluation for each population. Second, it is possible that depending on the roles of the students in the team, certain team members spend more time together and thus would have a better understanding of that person's personality over time. This could happen if the team divides tasks and assigns them to the same set of two or three student within the team. It is

also possible one team member is not pulling his or her own weight, and therefore is not interacting as often with the other team members.

8.3 Research Question 3

RQ3: Will student peer evaluations converge to match student self-evaluations?

Hypothesis 3: During the duration of a project, student peer evaluations will converge to student self-evaluations.

The results for this research question are also inconclusive. To test for agreement between the self and mean peer evaluations, a paired samples test was run between the self and mean peer evaluations at each iteration. In general, if the first iteration is ignored, the agreement behaves as expected increasing between Iteration 2 and Iteration 4. This shows that it is possible over time for peers to evaluate their teammates however at this time there is not enough evidence to claim that this would be the case every time. This could be due to the amount of time the students spent working on their project together compared to how often they collaborated online. Depending on the breakdown of the students in the meetings it is also possible that subsets of students got to know each other better because of the amount of time they spent working together. Similarly, the duration of the project can play a role in the interactions of the students. Knowing that the project ends at the end of the semester and the students might not have to work with their teammates in the future, it is possible that less effort was put into getting to know one another.

Like RQ1 and RQ2, there are some limitations of the results to consider. First, the tests are limited to the personalities of the engineering population at Clemson University,

especially within Mechanical Engineering. It has been shown that people of the same or similar personalities are better able to understand one another, and this could affect the results when comparing mean peer evaluations to self-evaluations. Additionally, these results can be limited by the number of teams and their compositions including how the teams were selected. Random or semi random selection does not always provide a balanced team in terms of experience. In ME4010 it is also possible that the team selection in place does not work for all teams, or the students did not fill out the initial team formation survey which led to problems later in the project. These factors could be taken into consideration in subsequent work with student engineering design teams.

Chapter Nine
FUTURE WORK

There are many different factors that can have an effect on an engineering design team. From gender, personality, and experience to the team selection method and goal of the project at hand. By understanding how peers view each other, steps can be taken to identify which combinations of personality work well together and which should be avoided for the purpose of the team's goal. This method goes beyond identifying only self-evaluations and allows researchers to take into account perception of peer's personalities, which can be important during the teaming process. A series of research questions have been developed to guide the direction of future work in this area.

RQ1. Do the hours the team has in contact with one another effect their self and peer evaluation convergence?

RQ2. Does the project duration effect the convergence of self and peer evaluations?

RQ3. What are the implications of forming teams based on profiles using the Five Factor Model?

For all three RQ's, the motivation stems from the limitations of the current study. Contact hours on a project are important not only for the successful completion of the project, but to ensure the team is performing their best. By tracking how long or how often students work together or meet either as a whole team or smaller subsets of people, there is potential for the convergence of the evaluations to be associated.

Similarly, the duration of the project might affect how the students act during the project. For longer projects there may be more motivation to understand one's teammates,

knowing that one would have to work with them for the given amount of time. In addition, if students know that there is no potential to work with the same people on a subsequent project, they may be less motivated to try and understand how they work well together.

Finally, it may be possible to identify personality "profiles" within populations of students. By using a profile, it can be determined whether or not groups of students with similar personalities are better or worse at understanding one another. This could also be used to identify the "worst" rater in a team and comparing these people to look for trends in the population. Using a profile might also make the Five Factor Model more appealing for use in engineering industry, due to the similarity with the Myers Briggs Personality Type Indicator. Profiles would also allow for factors such as gender, professional experience, or personal experience with teammates to be taken into account.

There are many characteristics of a team that can be identified during the team formation process. Personality is just one of numerous factors that have the potential to help a team be successful if taken into account. By controlling different factors in the team formation process and looking for patterns, the hope is that recommendations can be made to help teams in the future be successful.

APPENDICES

Appendix A - 50 Item Version of Five Factor Markers

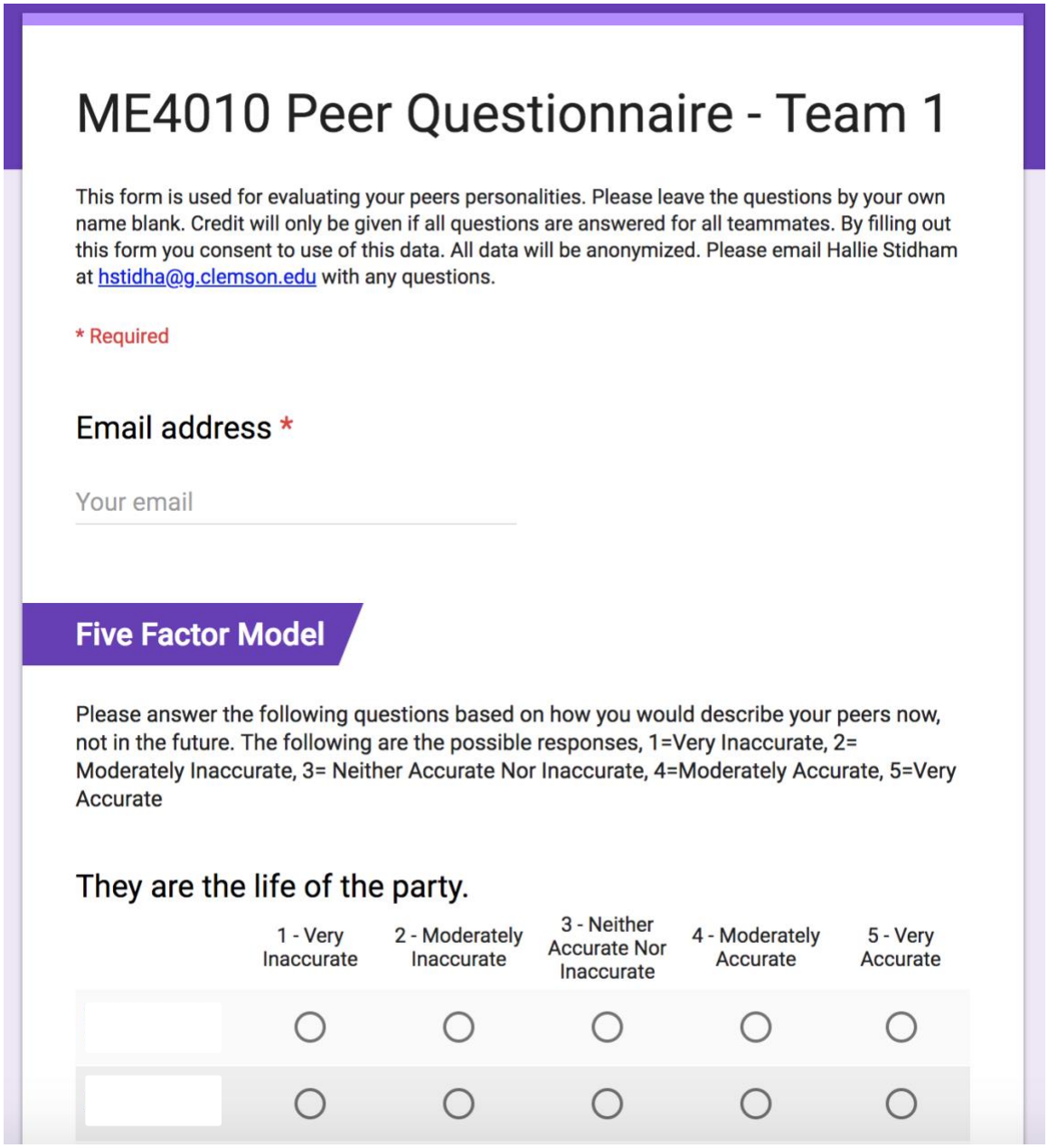
1: Extraversion, 2: Agreeableness, 3: Conscientiousness, 4: Neuroticism, 5: Openness
Keyed: +/-

Question	Factor	Key
1. Am the life of the party.	1	+
2. Feel little concern for others.	2	-
3. Am always prepared.	3	+
4. Get stressed out easily.	4	-
5. Have a rich vocabulary.	5	+
6. Don't talk a lot.	1	-
7. Am interested in people.	2	+
8. Leave my belongings around.	3	-
9. Am relaxed most of the time.	4	+
10. Have difficulty understanding abstract ideas.	5	-
11. Feel comfortable around people.	1	+
12. Insult people.	2	-
13. Pay attention to details.	3	+
14. Worry about things.	4	-
15. Have a vivid imagination.	5	+

16. Keep in the background.	1	-
17. Sympathize with other' feelings.	2	+
18. Make a mess of things.	3	-
19. Seldom feel blue.	4	+
20. Am not interested in abstract ideas.	5	-
21. Start conversations.	1	+
22. Am not interested in other people's problems.	2	-
23. Get chores done right away.	3	+
24. Am easily disturbed.	4	-
25. Have excellent ideas.	5	+
26. Have little to say.	1	-
27. Have a soft heart.	2	+
28. Often forget to put things back in their proper place.	3	-
29. Get upset easily.	4	-
30. Do not have a good imagination.	5	-
31. Talk to a lot of different people at parties.	1	+
32. Am not really interested in others.	2	-

33. Like order.	3	+
34. Change my mood a lot	4	-
35. Am quick to understand things.	5	+
36. Don’t like to draw attention to myself.	1	-
37. Take time out for others.	2	+
38. Shirk my duties.	3	-
39. Have frequent mood swings.	4	-
40. Use difficult words.	5	+
41. Don’t mind being the center of attention.	1	+
42. Feel others’ emotions.	2	+
43. Follow a schedule.	3	+
44. Get irritated easily.	4	-
45. Spend time reflecting on things.	5	+
46. Am quiet around strangers	1	-
47. Make people feel at ease.	2	+
48. Am exacting in my work.	3	+
49. Often feel blue.	4	-
50. Am full of ideas.	5	+

Appendix B - Sample Peer Evaluation with Accompanying IRB Agreement Statement



ME4010 Peer Questionnaire - Team 1

This form is used for evaluating your peers personalities. Please leave the questions by your own name blank. Credit will only be given if all questions are answered for all teammates. By filling out this form you consent to use of this data. All data will be anonymized. Please email Hallie Stidham at hstidha@g.clemson.edu with any questions.

* Required

Email address *

Your email

Five Factor Model

Please answer the following questions based on how you would describe your peers now, not in the future. The following are the possible responses, 1=Very Inaccurate, 2= Moderately Inaccurate, 3= Neither Accurate Nor Inaccurate, 4=Moderately Accurate, 5=Very Accurate

They are the life of the party.

	1 - Very Inaccurate	2 - Moderately Inaccurate	3 - Neither Accurate Nor Inaccurate	4 - Moderately Accurate	5 - Very Accurate
	○	○	○	○	○
	○	○	○	○	○

Appendix C - IBM SPSS Statistics Script for Statistical Tests

```
DO IF (Rater=1).
RECODE Rater1Time1 Rater1Time2 Rater1Time3 Rater1Time4 (ELSE=SYSMIS).
END IF.

DO IF (Rater=2).
RECODE Rater2Time1 Rater2Time2 Rater2Time3 Rater2Time4 (ELSE=SYSMIS).
END IF.

DO IF (Rater=3).
RECODE Rater3Time1 Rater3Time2 Rater3Time3 Rater3Time4(ELSE=SYSMIS).
END IF.

DO IF (Rater=4).
RECODE Rater4Time1 Rater4Time2 Rater4Time3 Rater4Time4(ELSE=SYSMIS).
END IF.

DO IF (Rater=5).
RECODE Rater5Time1 Rater5Time2 Rater5Time3 Rater5Time4(ELSE=SYSMIS).
END IF.

DO IF (Rater=6).
RECODE Rater6Time1 Rater6Time2 Rater6Time3 Rater6Time4(ELSE=SYSMIS).
END IF.
EXECUTE.


COMPUTE obs_openT1 = var(Rater1Time1 TO Rater6Time1).
COMPUTE rwg_openT1_un = 1-(obs_openT1/2).
COMPUTE rwg_openT1_ss = 1-(obs_openT1/1.34).
EXECUTE.

COMPUTE obs_openT2 = var(Rater1Time2 TO Rater6Time2).
COMPUTE rwg_openT2_un = 1-(obs_openT2/2).
COMPUTE rwg_openT2_ss = 1-(obs_openT2/1.34).
EXECUTE.

COMPUTE obs_openT3 = var(Rater1Time3 TO Rater6Time3).
COMPUTE rwg_openT3_un = 1-(obs_openT3/2).
COMPUTE rwg_openT3_ss = 1-(obs_openT3/1.34).
EXECUTE.
```

```
COMPUTE obs_openT4 = var(Rater1Time4 TO Rater6Time4).
COMPUTE rwg_openT4_un = 1-(obs_openT4/2).
COMPUTE rwg_openT4_ss = 1-(obs_openT4/1.34).
EXECUTE.


GLM rwg_openT1_ss rwg_openT2_ss rwg_openT3_ss rwg_openT4_ss
  /WSFACTOR=RWG 4 Polynomial
  /METHOD=SSTYPE(3)
  /PLOT=PROFILE(RWG)
  /EMMEANS=TABLES(RWG)
  /PRINT=DESCRIPTIVE
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=RWG.

GLM rwg_openT1_un rwg_openT2_un rwg_openT3_un rwg_openT4_un
  /WSFACTOR=RWG 4 Polynomial
  /METHOD=SSTYPE(3)
  /PLOT=PROFILE(RWG)
  /EMMEANS=TABLES(RWG)
  /PRINT=DESCRIPTIVE
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=RWG.

COMPUTE T1RMean=MEAN(Rater1Time1 TO Rater6Time1).
COMPUTE T2RMean=MEAN(Rater1Time2 TO Rater6Time2).
COMPUTE T3RMean=MEAN(Rater1Time3 TO Rater6Time3).
COMPUTE T4RMean=MEAN(Rater1Time4 TO Rater6Time4).
EXECUTE.

T-TEST PAIRS=SelfTime1 SelfTime2 SelfTime3 SelfTime4 WITH T1RMean
T2RMean T3RMean T4RMean (PAIRED)
  /CRITERIA=CI(.9500)
  /MISSING=ANALYSIS.

GLM rwg_openT1_ss rwg_openT2_ss rwg_openT3_ss rwg_openT4_ss
  /WSFACTOR=openness 4 Repeated
  /METHOD=SSTYPE(3)
  /PLOT=PROFILE(openness)
  /EMMEANS=TABLES(OVERALL)
  /PRINT=DESCRIPTIVE
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=openness.
```

Appendix D - Numerical R_{wg} Values for Both Populations

Table 9.1: Creative Inquiry R_{wg} Slightly Skewed Values

Iteration	O	C	E	A	N
IT1	0.8621	0.8147	0.8185	0.7998	0.7795
IT2	0.843	0.8035	0.7467	0.8105	0.7841
IT3	0.7343	0.7119	0.7163	0.6664	0.8374
IT4	0.3516	0.4115	0.31	0.549	0.4675

Table 9.2: Creative Inquiry R_{wg} Uniform Values

Iteration	O	C	E	A	N
IT1	0.9076	0.8759	0.8784	0.8658	0.8523
IT2	0.8948	0.8683	0.8303	0.873	0.8553
IT3	0.822	0.8087	0.8099	0.7765	0.8911
IT4	0.5656	0.6057	0.5377	0.6978	0.6432

Table 9.3: ME4010 R_{wg} Slightly Skewed Values

Iteration	O	C	E	A	N
IT1	0.884	0.8338	0.8286	0.846	0.782
IT2	0.872	0.7925	0.8096	0.846	0.751
IT3	0.832	0.7723	0.7855	0.817	0.679
IT4	0.802	0.7506	0.7578	0.799	0.738

Table 9.4: ME4010 R_{wg} Uniform Values

Iteration	O	C	E	A	N
IT1	0.922	0.8886	0.8851	0.8969	0.8542
IT2	0.9145	0.861	0.8725	0.8968	0.8332
IT3	0.8872	0.8474	0.8563	0.8776	0.7848
IT4	0.8675	0.8329	0.8377	0.8655	0.8242

<u>Appendix E</u> - Uniform R_{wg} values for Peer-Evaluations in Creative Inquiry Over Time

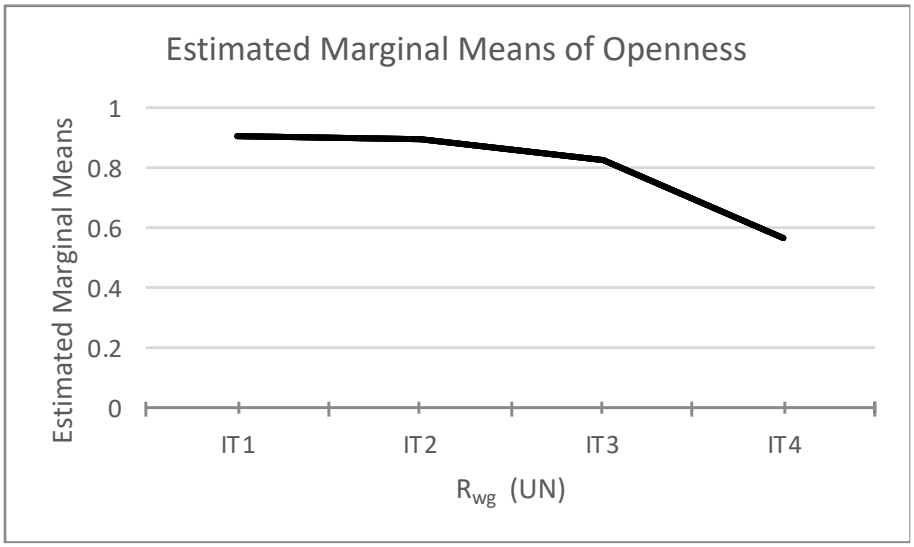


Figure 9.1: CI R_{wg} (UN) Values for Openness Over Time

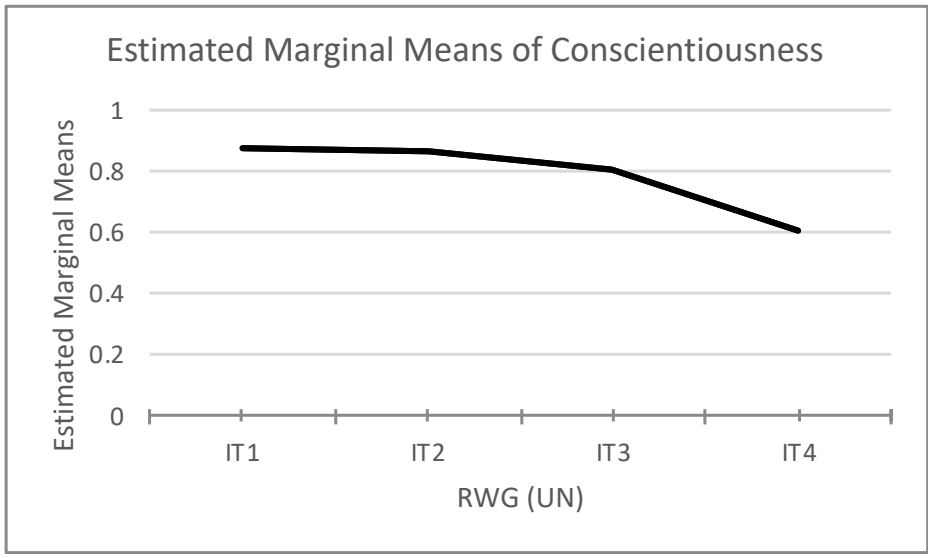


Figure 9.2: CI R_{wg} (UN) Values for Conscientiousness Over Time

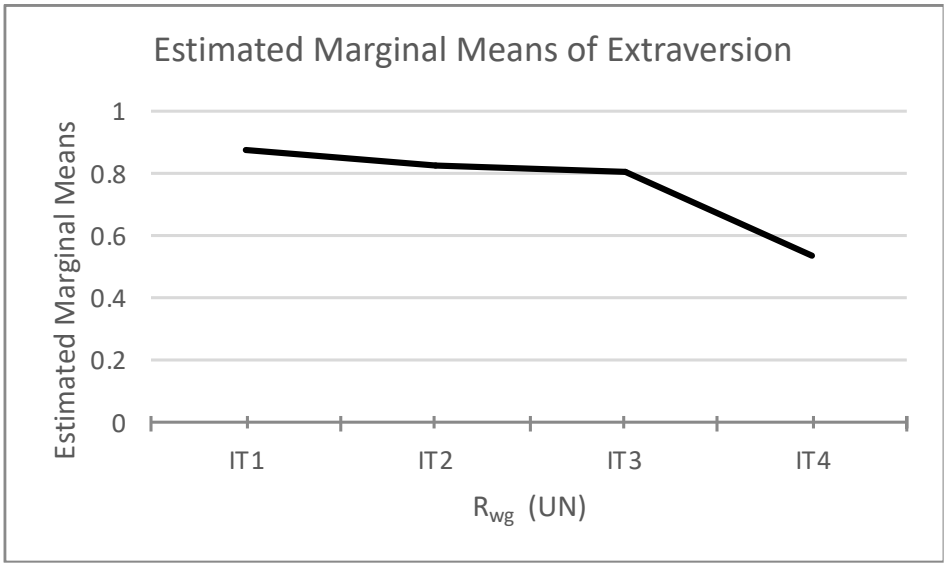


Figure 9.3: CI R_{wg} (UN) Values for Extraversion Over Time

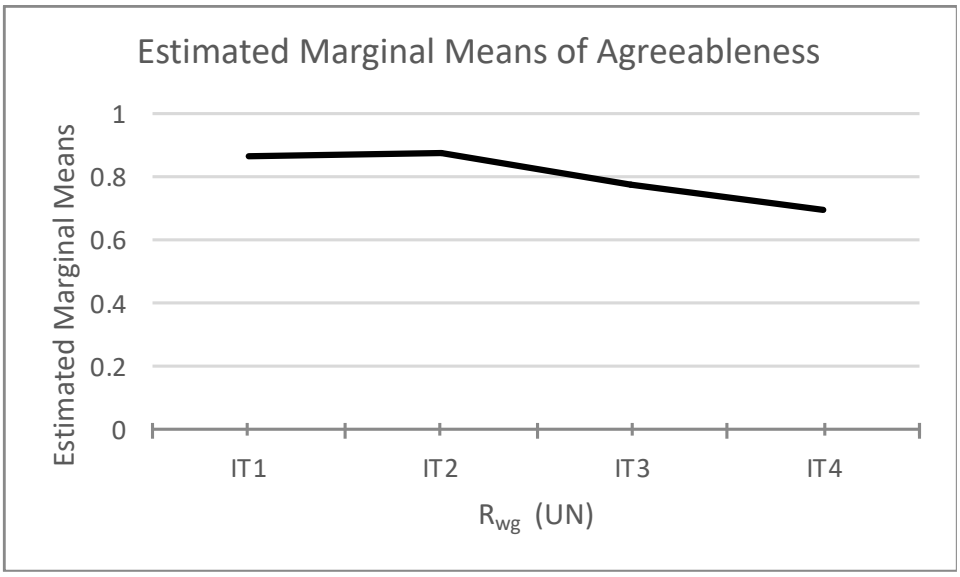


Figure 9.4: CI R_{wg} (UN) Values for Agreeableness Over Time

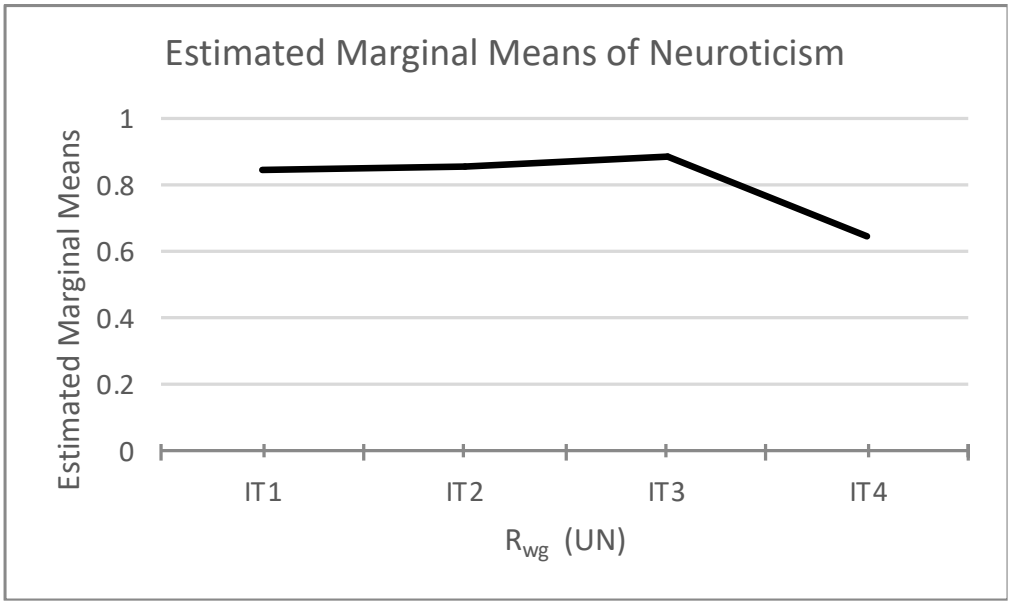


Figure 9.5: CI R_{wg} (UN) Values for Neuroticism Over Time

Appendix F - Uniform r_{wg} values for Peer-Evaluations in ME4010 Over Time

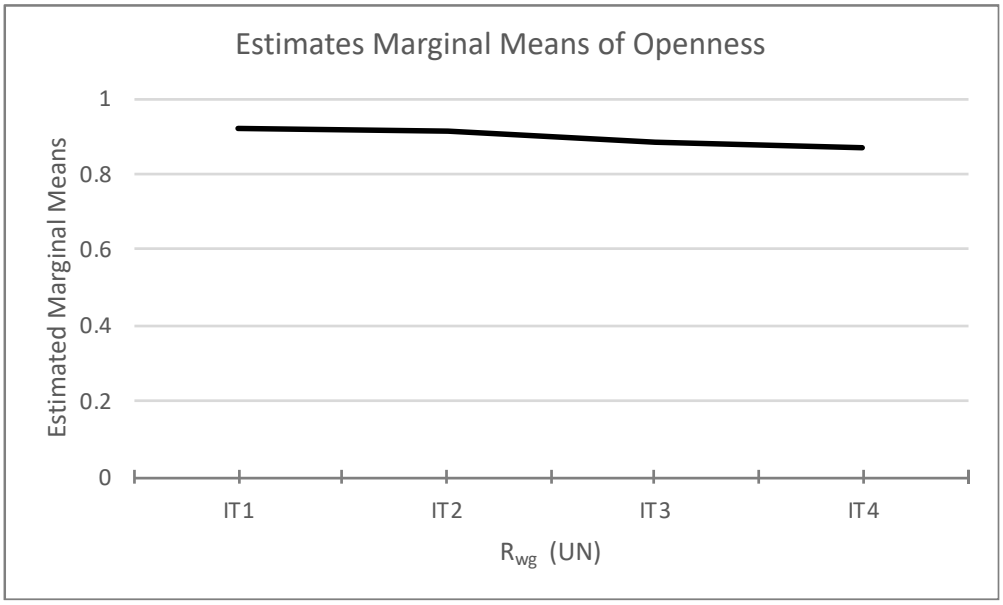


Figure 9.6: ME4010 R_{wg} (UN) Values for Openness Over Time

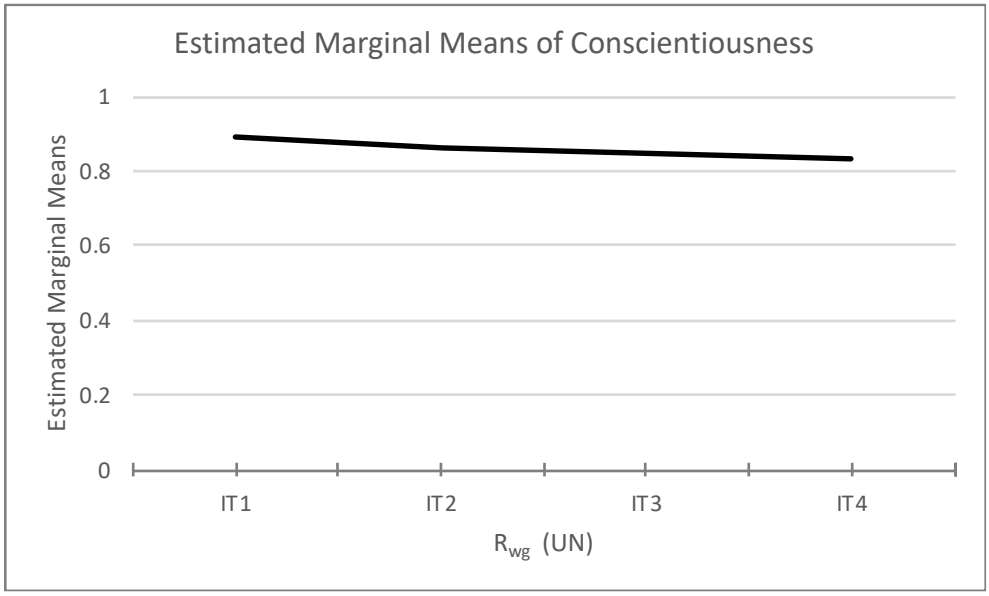


Figure 9.7: ME4010 R_{wg} (UN) Values for Conscientiousness Over Time

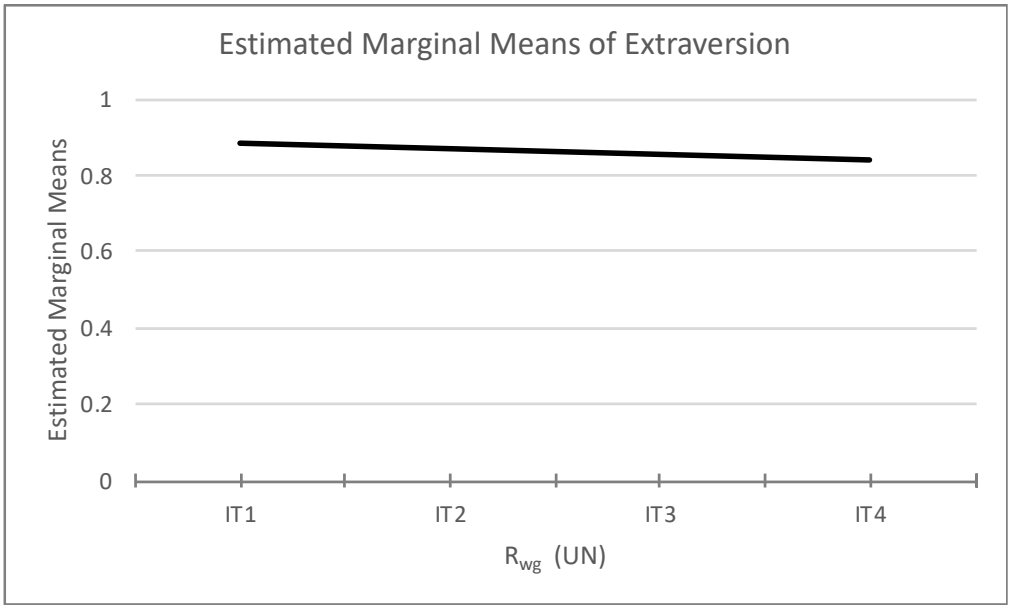


Figure 9.8: ME4010 R_{wg} (UN) Values for Extraversion Over Time

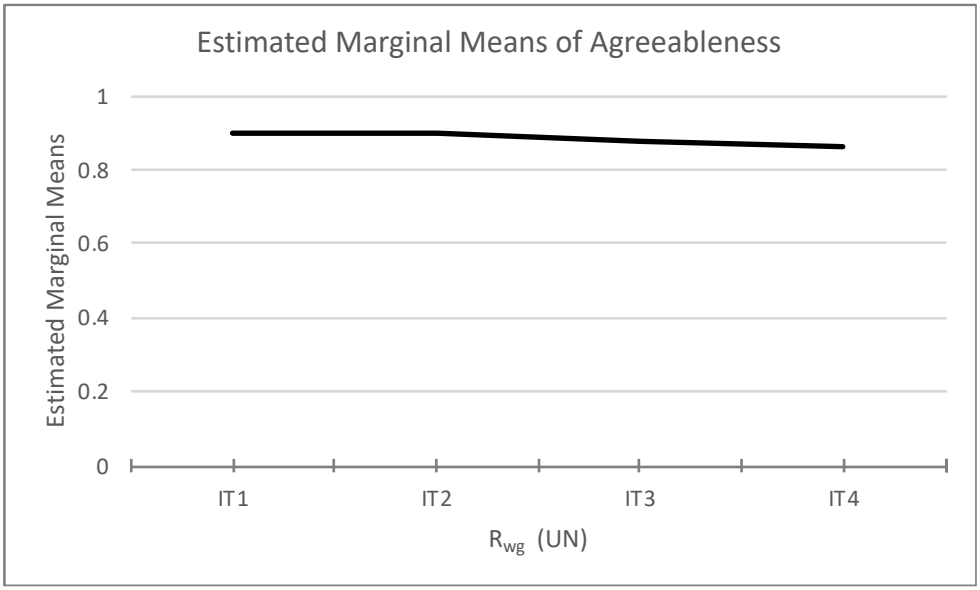


Figure 9.9: ME4010 R_{wg} (UN) Values for Agreeableness Over Time

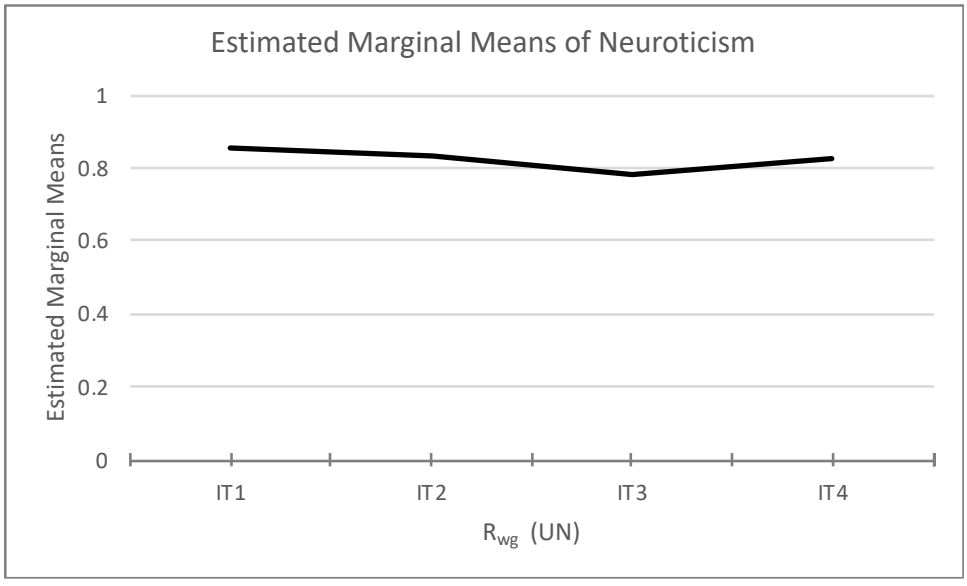


Figure 9.10: ME4010 R_{wg} (UN) Values for Neuroticism Over Time

REFERENCES

André, M., Baldoquín, M. G., & Acuña, S. T. (2011). Formal model for assigning human resources to teams in software projects. *Information and Software Technology*, *53*(3), 259-275.

Bacon, D. R., Stewart, K. A., & Silver, W. S. (1999). Lessons from the best and worst student team experiences: How a teacher can make the difference. *Journal of Management Education*, *23*(5), 467-488.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. Personnel psychology, 44(1), 1-26.

Berthelsdorf, F. A., & Stone, R. B. (2017, August). Creativity of Concept Ideation Methods As Affected by Team Personality. In *ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. V007T06A022-V007T06A022). American Society of Mechanical Engineers.

Bjork, R., & Druckman, D. (1992). *In the mind's eye: Enhancing human performance*. National Research Council. National Academies Press.

Borrego, M., Karlin, J., McNair, L. D., & Beddoes, K. (2013). Team effectiveness theory from industrial and organizational psychology applied to engineering student project teams: A research review. *Journal of Engineering Education*, *102*(4), 472-512.

Boyle, G. J. (1995). Myers-Briggs Type Indicator (MBTI): Some Psychometric Limitations. *Australian Psychologist*, *30*(1), 71-74.

Burton, L. J., & Dowling, D. G. (2010). The effects of gender on the success of a cohort of engineering students. In *Proceedings of Engineering Education Conference (EE 2010)*(pp. 1-10). Higher Education Academy Engineering Subject Centre.

Chen, C., & Messner, J. I. (2010). A recommended practices system for a global virtual engineering team. *Architectural Engineering and Design Management*, *6*(3), 207-221.

Chiocchio, F., Grenier, S., O'Neill, T. A., Savaria, K., & Willms, J. D. (2012). The effects of collaboration on performance: A multilevel validation in project teams. International Journal of Project Organisation and Management, 4(1), 1-37.

Clinebell, S., & Stecher, M. (2003). Teaching Teams to be Teams: An Exercise Using the Myers-Briggs® Type Indicator and the Five-Factor Personality Traits—. Journal of Management Education, 27(3), 362-383.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, *4*(1), 5.

Costa, P. T., & McCrae, R. R. (1997). Longitudinal stability of adult personality. In Handbook of personality psychology (pp. 269-290).

De Raad, B., & Schouwenburg, H. C. (1996). Personality in learning and education: A review. European Journal of personality, 10(5), 303-336.

De Vreede, T., de Vreede, G. J., Ashley, G., & Reiter-Palmon, R. (2012, January). Exploring the effects of personality on collaboration technology transition. In *System Science (HICSS), 2012 45th Hawaii International Conference on* (pp. 869-878). IEEE.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, *41*(1), 417-440.

Doolen, T. L., Hacker, M. E., & Van Aken, E. (2006). Managing organizational context for engineering team effectiveness. *Team Performance Management: An International Journal*, *12*(5/6), 138-154.

Duhe, S. (2009). What's Your Type? Using the Myers–Briggs Personality Inventory to Improve Team Performance. *Communication Teacher*, *23*(4), 142-147.

Dym, C. L., & Little, P (1999). *Engineering Design: A Project-Based Introduction*. John Wiley and Sons, New York, NY.

Ehrhart, K. H., Roesch, S. C., Ehrhart, M. G., & Kilian, B. (2008). A test of the factor structure equivalence of the 50-item IPIP Five-factor model measure across gender and ethnic groups. *Journal of Personality Assessment*, *90*(5), 507-516.

Fazelpour, M., Patel, A., Shankar, P., & Summers, J. D. (2017, August). A User Study on Exploring the Sequencing of Unit Cell Design Guidelines. In ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (pp. V007T06A027-V007T06A027). American Society of Mechanical Engineers.

Felder, R. M., Felder, G. N., & Dietz, E. J. (2002). The effects of personality type on engineering student performance and attitudes. *Journal of engineering education*, *91*(1), 3-17.

Feldt, R., Torkar, R., Angelis, L., & Samuelsson, M. (2008, May). Towards individualized software engineering: empirical studies should collect psychometrics. In *Proceedings of the 2008 international workshop on Cooperative and human aspects of software engineering* (pp. 49-52). ACM

Freiheit, T. (2014, August). Validity and Predictive Ability of Wilde's Cognitive Teamology Model to Design Team Outcomes. In *ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. V007T07A034-V007T07A034). American Society of Mechanical Engineers.

Furnham, A. (1996). The big five versus the big four: the relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality. *Personality and Individual Differences*, *21*(2), 303-307.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological assessment*, *4*(1), 26.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, *7*(1), 7-28.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality*, *40*(1), 84-96.

Goldberg, J. R. (2010). Developing teamwork skills in capstone design courses [Senior Design]. IEEE Engineering in Medicine and Biology Magazine, 29(2), 141-142.

Hannah, R., Joshi, S., & Summers, J. D. (2012). A user study of interpretability of engineering design representations. Journal of Engineering Design, 23(6), 443-468.

Hirsh, J. B., & Peterson, J. B. (2008). Predicting creativity and academic success with a "fake-proof" measure of the Big Five. *Journal of Research in Personality*, *42*(5), 1323-1333.

Holler, M., Uebernickel, F., & Brenner, W. (2017). Defining Archetypes of E-Collaboration for Product Development in The Automotive Industry.

Hox, J. J. (2010). Analyzing longitudinal data. In Multilevel Analysis: Techniques and applications (2nd Ed., pp. 79-111). New York, NY: Routledge.

Hughes, R. L., & Jones, S. K. (2011). Developing and assessing college student teamwork skills. New Directions for Institutional Research, 2011(149), 53-64.

Jackson, S. L., Parker, C. P., & Dipboye, R. L. (1996). A comparison of competing models underlying responses to the Myers-Briggs Type Indicator. *Journal of Career Assessment*, *4*(1), 99-111.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, *2*(1999), 102-138.

Joshi, S., & Summers, J. D. (2010, June). Investigating information loss in collaborative design: a case study with capstone design project. In National Capstone Conference 2010, Boulder, CO (No. 51).

Joshi, S., Morkos, B., & Summers, J. D. (2011, January). Mapping problem and requirements to final solution: A document analysis of capstone design projects. In ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (pp. 569-578). American Society of Mechanical Engineers.

Joshi, S., & Summers, J. D. (2015). Requirements evolution: Understanding the type of changes in the requirement document of novice designers. In ICoRD'15–Research into Design Across Boundaries Volume 2 (pp. 471-481). Springer, New Delhi.

Kanij, T., Merkel, R., & Grundy, J. (2015, May). An empirical investigation of personality traits of software testers. In *Proceedings of the Eighth International Workshop on Cooperative and Human Aspects of Software Engineering* (pp. 1-7). IEEE Press.

Karn, J., & Cowling, T. (2006, September). A follow up study of the effect of personality on the performance of software engineering teams. In *Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering* (pp. 232-241). ACM.

Kearney, E., Gebert, D., & Voelpel, S. C. (2009). When and how diversity benefits teams: The importance of team members' need for cognition. *Academy of Management journal*, *52*(3), 581-598.

Kichuk, S. L., & Wiesner, W. H. (1997). The big five personality factors and team performance: implications for selecting successful product design teams. *Journal of Engineering and Technology Management*, *14*(3-4), 195-221.

Kratzer, J., Leenders, R. T. A., & Van Engelen, J. M. (2010). The social network among engineering design teams and their creativity: A case study among teams in two product development programs. *International Journal of Project Management*, *28*(5), 428-436.

Kress, G., & Schar, M. (2011). Initial conditions: The structure and composition of effective design teams. In DS 68-7: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 7: Human Behaviour in Design, Lyngby/Copenhagen, Denmark, 15.-19.08. 2011.

Layton, R. A., Loughry, M. L., Ohland, M. W., & Ricco, G. D. (2010). Design and Validation of a Web-Based System for Assigning Members to Teams Using Instructor-Specified Criteria. *Advances in Engineering Education*, *2*(1), n1.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. Organizational research methods, 11(4), 815-852.

LePine, J. A., Buckman, B. R., Crawford, E. R., & Methot, J. R. (2011). A review of research on personality in teams: Accounting for pathways spanning levels of theory and analysis. Human Resource Management Review, 21(4), 311-330.

Lewis, P., Aldridge, D., & Swamidass, P. M. (1998). Assessing teaming skills acquisition on undergraduate project teams. *Journal of Engineering Education*, *87*(2), 149-155.

Licorish, S., Philpott, A., & MacDonell, S. G. (2009, May). Supporting agile team composition: A prototype tool for identifying personality (In) compatibilities. In *Proceedings of the 2009 ICSE Workshop on Cooperative and Human Aspects on Software Engineering* (pp. 66-73). IEEE Computer Society.

Lilikovych, S., & Sokol, V. A Human-Centered Approach to Assessment of Team Diversity in Multi-Version Software Development.

Lingard, R., & Barkataki, S. (2011, October). Teaching teamwork in engineering and computer science. In *Frontiers in Education Conference (FIE), 2011* (pp. F1C-1). IEEE.

Lounsbury, J. W., Sundstrom, E., Loveland, J. M., & Gibson, L. W. (2003). Intelligence,"Big Five" personality traits, and work drive as predictors of course grade. *Personality and individual differences*, *35*(6), 1231-1239.

Lugo, J. E., Zapata-Ramos, M. L., & Puig, C. P. (2017, August). Exploration of Entrepreneurial Student Teams Performance and Student Team Members' Personality via the Big Five Test. In *ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. V007T06A021-V007T06A021). American Society of Mechanical Engineers.

Magness, J. B., & Roslewicz, E. A. (2009, October). An innovative Myers-Briggs type indicator educational experience used in an information technology capstone course. In *Proceedings of the 10th ACM conference on SIG-information technology education* (pp. 144-149). ACM.

May, A., & Carter, C. (2001). A case study of virtual team working in the European automotive industry. *International Journal of Industrial Ergonomics*, *27*(3), 171-186.

McCarley, N. G., & Carskadon, T. G. (1983). Test-retest reliabilities of scales and subscales of the Myers-Briggs Type Indicator and of criteria for clinical interpretive hypotheses involving them. *Research in Psychological Type*, *6*, 24-36.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, *52*(1), 81.

McCrae, R. R., & Costa, P. T. (1989). Reinterpreting the Myers-Briggs type indicator from the perspective of the five-factor model of personality. *Journal of personality*, *57*(1), 17-40.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, *60*(2), 175-215.

McCrae, R. R., & Costa Jr, P. T. (1997). Personality trait structure as a human universal. *American psychologist*, *52*(5), 509.

McNair, L. D., Newswander, C., Boden, D., & Borrego, M. (2011). Student and faculty interdisciplinary identities in self-managed teams. *Journal of Engineering Education*, *100*(2), 374-396.

Meyer, R. D., Mumford, T. V., Burrus, C. J., Campion, M. A., & James, L. R. (2014). Selecting null distributions when calculating rwg: A tutorial and review. *Organizational Research Methods*, *17*(3), 324-345.

Montequín, V. R., Fernández, J. M., Balsera, J. V., & Nieto, A. G. (2013). Using MBTI for the success assessment of engineering teams in project-based learning. *International journal of technology and design education*, *23*(4), 1127-1146.

Morkos, B., & Summers, J. D. (2010). Implementing design tools in capstone design projects: requirements elicitation through use of personas. In National Capstone Conf (p. 50).

Morkos, B., & Summers, J. D. (2013). A study of designer familiarity with product and user during requirement elicitation. International Journal of Computer Aided Engineering and Technology 8, 5(2-3), 139-158.

Morkos, B., Summers, J. D., & Thoe, S. (2014). A comparative survey of domestic and international experiences in capstone design. International Journal of Engineering Education

Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator* (Vol. 3). Palo Alto, CA: Consulting Psychologists Press.

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual differences*, *43*(5), 971-990.

Ogot, M., & Okudan, G. E. (2006). The Five-Factor Model personality assessment for improved student design team performance. *European Journal of Engineering Education*, *31*(5), 517-529.

Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C. J., ... & Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self-and peer evaluation. *Academy of Management Learning & Education*, *11*(4), 609-630.

Okudan, G. E., Schmidt, L. C., Hernandez, N. V., Jablokow, K., & Lin, C. Y. (2012, August). Questioning the Impact of Personality on Design Outcomes: Should We Take It Into Account?. In *ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. 95-102). American Society of Mechanical Engineers.

Ostergaard, K. J., & Summers, J. D. (2009). Development of a systematic classification and taxonomy of collaborative design activities. *Journal of Engineering Design*, *20*(1), 57-81.

Pahl, G., Beitz, W., Schulz, H. J., & Jarecki, U. (1996). *Engineering Design: A Systematic Approach*, Springer Verlag.

Paretti, M., Layton, R., Laguette, S., & Speegle, G. (2011). Managing and mentoring capstone design teams: Considerations and practices for faculty. International Journal of Engineering Education, 27(6), 1192.

Phelan, K. T., Pearce, B., Summers, J. D., & Kurz, M. E. (2016). Evaluating configuration rule implementation through a graph-based user interface. Tools and Methods for Competitive Engineering, Aix-en-Provence, France, Paper, (69).

Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research*, *57*(3), 210.

Reilly, R. R., Lynn, G. S., & Aronson, Z. H. (2002). The role of personality in new product development team performance. *Journal of Engineering and Technology Management*, *19*(1), 39-58.

Shen, S. T., Prior, S. D., White, A. S., & Karamanoglu, M. (2007). Using personality type differences to form engineering design teams. *Engineering education*, *2*(2), 54-66.

Smith, K. A. (1995, November). Cooperative learning: Effective teamwork for engineering classrooms. In *Frontiers in Education Conference, 1995. Proceedings., 1995* (Vol. 1, pp. 2b5-13). IEEE.

Summers, J. D., Joshi, S., & Morkos, B. (2014, August). Requirements evolution: Relating functional and non-functional requirement change on student project success. In ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (pp. V003T04A002-V003T04A002). American Society of Mechanical Engineers.

Takai, S., & Esterman, M. (2017, August). Towards a Better Design Team Formation: A Review of Team Effectiveness Models and Possible Measurements of Design-Team Inputs, Processes, and Outputs. In *ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. V003T04A018-V003T04A018). American Society of Mechanical Engineers.

Teegavarapu, S., Summers, J. D., & Mocko, G. M. (2008, January). Case study method for design research: A justification. In *ASME 2008 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. 495-503). American Society of Mechanical Engineers.

Thimmaiah, S., Phelan, K., & Summers, J. D. (2017). An Experimental Study on the Influence That Failure Number, Specialization, and Controls Have on Confidence in Predicting System Failures. Journal of Mechanical Design, 139(1), 011102.

Toh, C. A., & Miller, S. R. (2016). Creativity in design teams: the influence of personality traits and risk attitudes on creative concept selection. *Research in Engineering Design*, *27*(1), 73-89.

Ullman, D. G. (2010). *The Mechanical Design Process*, McGraw-Hill Science/Engineering/Math.

Varvel, T., Adams, S. G., Pridie, S. J., & Ruiz Ulloa, B. C. (2004). Team effectiveness and individual Myers-Briggs personality dimensions. *Journal of management in Engineering*, *20*(4), 141-146.

Verhoeven, D. C. (2016). An examination of trust, distrust, and shared leadership in distributed teams (Doctoral dissertation, Clemson University).

Wilde, D. J. (2008). *Teamology: the construction and organization of effective teams*. Springer Science & Business Media.

Yilmaz, S., Daly, S. R., Christian, J. L., Seifert, C. M., & Gonzalez, R. (2014). Can experienced designers learn from new tools? A case study of idea generation in a professional engineering team. *International Journal of Design Creativity and Innovation*, *2*(2), 82-96.

Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident Analysis & Prevention*, *42*(5), 1517-1522.